# Evolutionary OT and the Emergence of Possession Splits

Workshop “Logic, Neural Networks, and Optimality Theory”
ZAS Berlin
July 24, 2003

***Gerhard Jäger***
University of Potsdam
jaeger@ling.uni-potsdam.de
www.ling.uni-potsdam.de/~jaeger/

***Anette Rosenbach***
University of Düsseldorf
ar@phil-fak.uni-duesseldorf.de

# Overview

1. The phenomenon of possession splits
2. Harmonic alignment and Stochastic OT
3. The Iterated Learning Model of Language Evolution
4. The statistical patterns of possessive constructions in spoken English
5. Simulating the emergence of harmonic alignment via iterated learning
6. Conclusion

# 2.1. Aissen and Bresnan: Harmonic Alignment

- structural scale (nominal scale):

$$\mathrm{SPEC} > \textsc{non}\text{-}\mathrm{SPEC}$$

*(basically: prenominal position is more prominent than post-nominal position*

- substantive markedness scale
  - animacy hierarchy

**human** > **animate** > **inanimate**

  - definiteness scale

**pronoun** > **proper N** > **def** > **indef**

- harmonic alignment of structural/nominal and substantive scales
- leads to two universal sub-hierarchies

$$*SPEC/inanimate \gg *SPEC/animate \gg *SPEC/human$$

$$*non\text{-}SPEC/inanimate \gg *non\text{-}SPEC/animate \gg *non\text{-}SPEC/human$$

- ignoring category “human” for simplicity gives

$$*SPEC/inanimate \gg *SPEC/animate$$

$$*non\text{-}SPEC/animate \gg *non\text{-}SPEC/inanimate$$

- six possible rankings (respecting the universal sub-hierarchies above)

1. *Spec/inanim $\gg$ *Spec/anim $\gg$ *NSpec/anim $\gg$ *NSpec/inanim
2. *Spec/inanim $\gg$ *NSpec/anim $\gg$ *Spec/anim $\gg$ *NSpec/inanim
3. *Spec/inanim $\gg$ *NSpec/anim $\gg$ *NSpec/inanim $\gg$ *Spec/anim
4. *NSpec/anim $\gg$ *Spec/inanim $\gg$ *Spec/anim $\gg$ *NSpec/inanim
5. *NSpec/anim $\gg$ *Spec/inanim $\gg$ *NSpec/inanim$\gg$ *Spec/anim
6. *NSpec/anim $\gg$ *NSpec/inanim$\gg$ *Spec/inanim $\gg$ *Spec/anim

- predicts three language types

  A. all possessors are realized postnominally (ranking 1)

  B. animate possessors are prenominal, inanimates one postnominal (ranking 2–5)

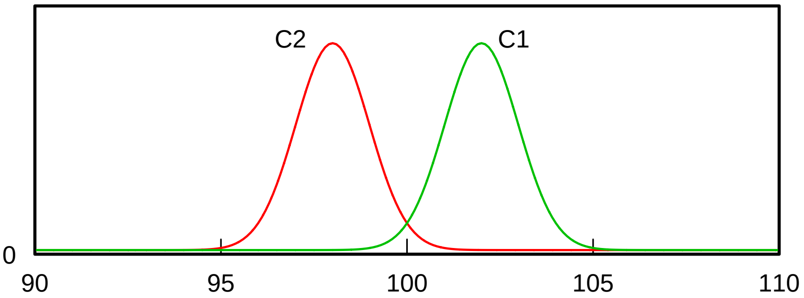
  C. all possessors are realized prenominally (ranking 6)

- implicative universal:
  **If possessors of a substantive category $C$ are realized prenominally, then all possessors of a more prominent category are also realized prenominally.**

# 2.2. Stochastic Optimality Theory (StOT)

- probabilistic grammar
- assigns probability distribution over possible meanings for a given form (and vice versa)
- Two modifications of standard OT (cf. Boersma 1998)

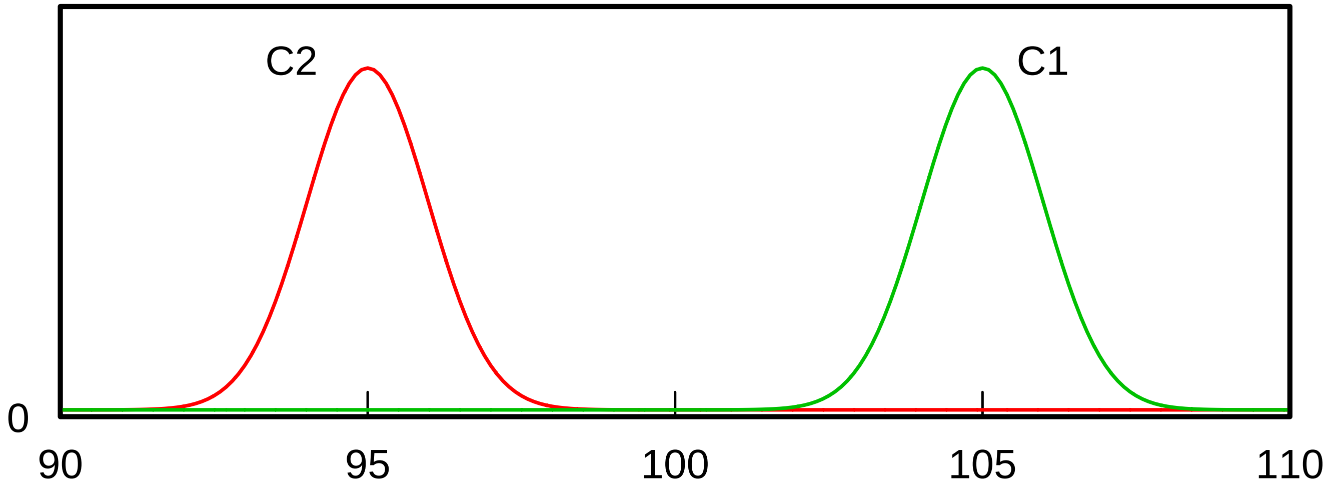
1. **constraint ranking on a continuous scale** distance between constraints matters
2. **stochastic evaluation** actual ordering of constraints varies, with probabilities depending on continuous ranking



- Absolute size of the distance between conflicting constraints determines their interaction:
  - difference between mean values $> 10$ units:

C$_1$ dominates C$_2$ categorically

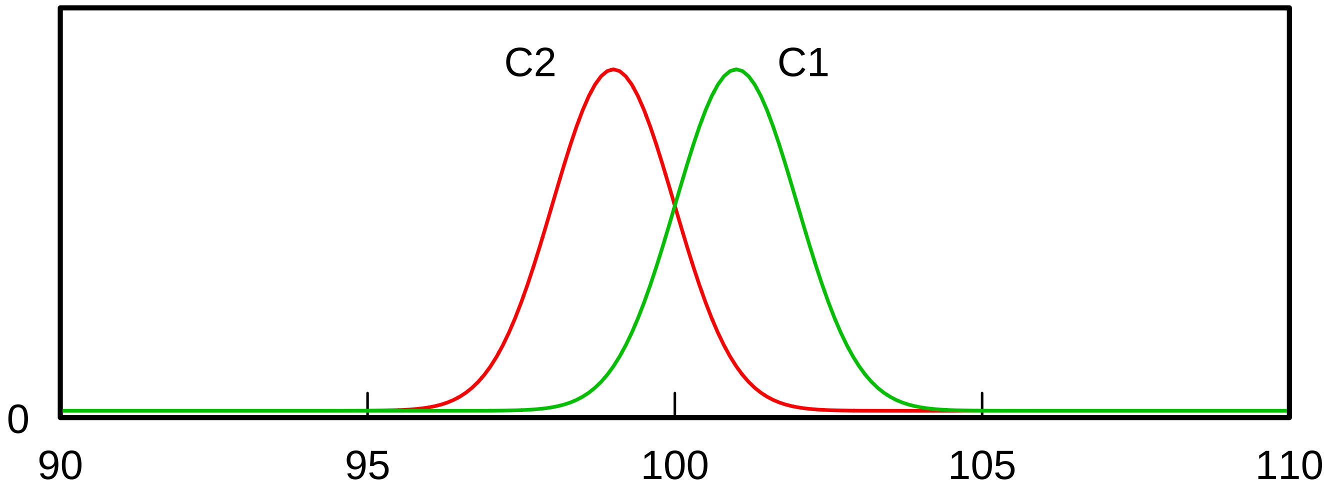
$$p(C_2 > C_1) < 10^{-10}$$



- difference $\approx 2$:

preference for obeying $C_1$, but obeying $C_2$ is still grammatical

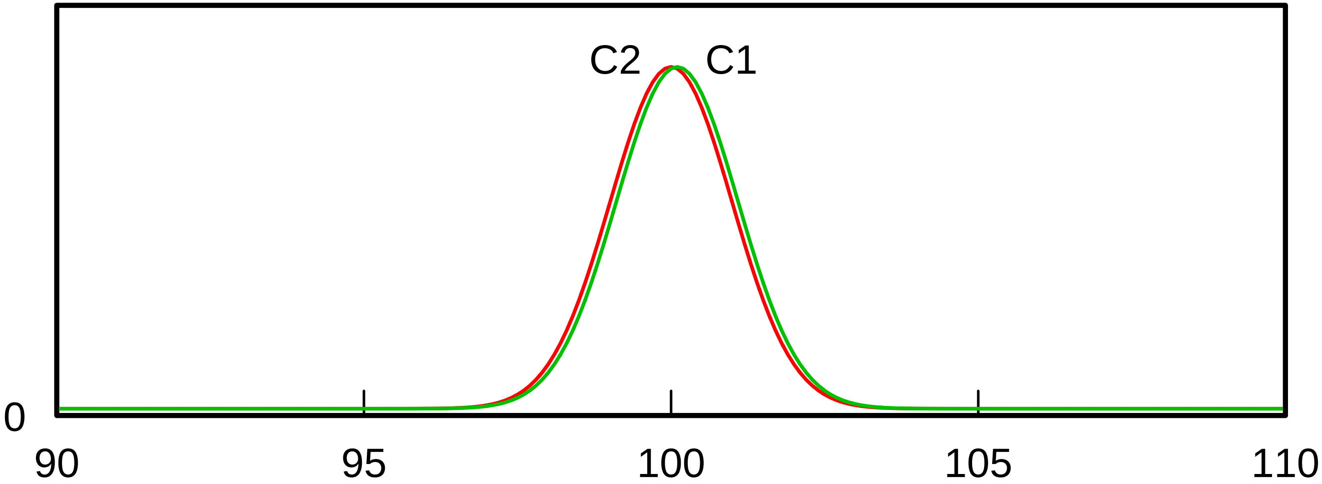
$$p(C_2 > C_1) \approx 30\%$$



- Both constraints are roughly equally ranked:

free variation

$$p(C_2 > C_1) = 50\%$$



## 2.3. Stochastic reinterpretation of harmonic alignment

- “universal sub-hierarchies” do not exist in StOT
- every constraint can outrank any other constraint with a positive probability
- stochastic interpretation of sub-hierarchies:
  - *C1 $\gg$ C2 universally* means:
  - *In each language, the average rank of C1 is higher than the average rank of C2.*
  - In other words:
    *In all languages, C1 $\gg$ C2 is more likely than C2 $\gg$ C1*
- harmonic alignment for possessor realization boils down to down to:

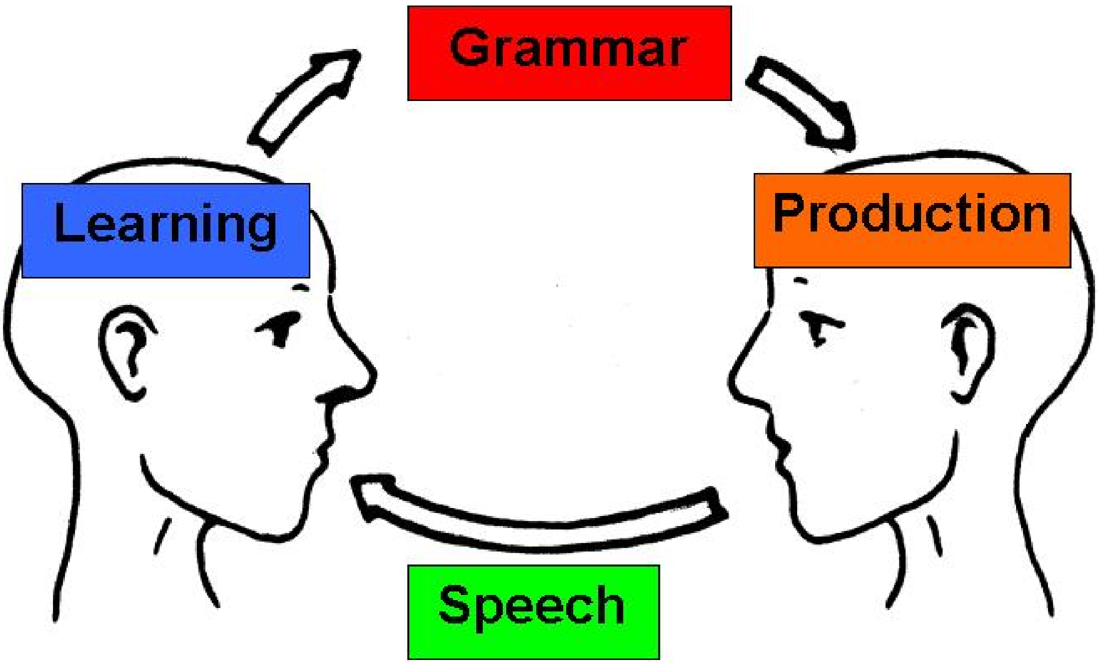
$$P(\text{Spec}|\text{human}) > P(\text{Spec}|\text{anim}) > P(\text{Spec}|\text{inanim})$$

$$P(\text{Spec}|\text{pron}) > P(\text{Spec}|\text{defNP}) > P(\text{Spec}|\text{indefNP})$$

## 2.4. StOT and iterated learning

- similarity between language and biological systems
  - grammar is self-replicating system (like genome)
  - replication (via language acquisition and language use) is subject to random variation
  - differential replicative success of competing variants
  - determined by differential adaptation to environment (i.e. learning and usage)

# 3.1. Iterated Learning



## 3.2. Filtered learning

- Kirby:
  - only successfully parsed observations have an effect on learning
  - parser sometimes fails
  - input for learning algorithm thus not raw performance data
  - parser acts as a filter
  - high parsing complexity lowers $\approx$ low impact on learning (and vice versa)

# 4.2. Filters

1. parsing complexity (in the sense of Hawkins' EIC metric)
    - favors uniform directionality of heads
    - VO-languages: prenominal poss. more complex than postnominal
    - OV-languages: prenominal poss. less complex than postnominal
2. semantic processing complexity
    - if possessive NP is definite:
        - possessor is referential anchor
        - referent of possessor must be processed **before** referent of possessive phrase as a whole is processed
        - prenominal position of possessor facilitates processing
    - does not apply to indefinite possessive NPs
    - cross-linguistically not parameterized

# 5.1. Learning bias

## VO-language pattern

| | GEN-N | N-GEN |
|---|---|---|
| *definite head* | pld ... pld ... pld ... pld | pld ... pld ... pld ... pld |
| *indefinite head* | pld ... pld ... pld ... pld | pld ... pld ... pld ... pld |

- constraint inventory not sensitive to definiteness of the head
- however, statistical correlation

definite head ≈ animate/definite possessor
indefinite head ≈ inanimate/indefinite possessor

- learning bias towards possession split

## 5.2. Bidirectional Gradual Learning Algorithm

- Gradual Learning Algorithm (Boersma):
  - gradually adjusts constraint rankings on basis of observations
  - converges towards a stochastic ranking that matches the probability distribution of the observed data
- Bidirectional Gradual Learning Algorithm (BiGLA):
  - variant of GLA
  - simultaneous production- and interpretation-oriented learning
  - converges towards stochastic constraint ranking that approximates the empirical conditional probability distribution P(form|meaning) and P(meaning|form)

## 5.3. Asymmetric Bidirectional Evaluation

- variant of Bidirectional OT
- Intuition: speaker tries to maximize his chance of getting his message across (first priority) while minimizing the constraint violation profile (second priority)
- related proposals by Boersma, Beaver, Vogel, ...

**Definition 1 (AB-optimality)**

- *A form-meaning pair* $\langle f, m\rangle$ *is* hearer-optimal *iff* $\langle f, m\rangle \in \mathbf{GEN}$ *and there is no alternative meaning* $m'$ *such that* $\langle f, m'\rangle \in \mathbf{GEN}$ *and* $\langle f, m'\rangle < \langle f, m\rangle$.
- *A form-meaning pair* $\langle f, m\rangle$ *is* optimal *iff either it is hearer-optimal and there is no alternative form* $f'$ *such that* $\langle f', m\rangle$ *is hearer-optimal and* $\langle f', m\rangle < \langle f, m\rangle$*, or there is no hearer-optimal* $\langle f', m\rangle$*, and there is no* $\langle f', m\rangle \in \mathbf{GEN}$ *such that* $\langle f', m\rangle < \langle f, m\rangle$.

# 5.4. The experiment

- Generator:
  - eight meanings (head: +/- definite, possessor: +/-definite, +/-animate)
  - three forms (possessor can be prenominal and postnominal, in the latter case the article can be definite or indefinite)
  - definiteness of head must be compatible with overt article (if present)
- Constraints:
  - eight alignment constraints
  - two interpretive constraints determining the definiteness of the head in the absence of an overt determiner
- Filter:
  - prenominal possessor: 2% are filtered out
  - definite head with postnominal possessor: 3% are filtered out

- Frequencies in 0th generation:
  - relative frequencies of meanings as found in ICE corpus
  - pre- and postnominal genitives equally probable

| head definite? | possessor definite? | possessor animate? | prenominal possessor | postnominal possessor | |
|---|---|---|---|---|---|
| | | | | *the* | *a* |
| yes | yes | yes | 1739 | 1739 | – |
| yes | yes | no | 210 | 210 | – |
| yes | no | yes | 47 | 47 | – |
| yes | no | no | 230 | 230 | – |
| no | yes | yes | 119 | – | 119 |
| no | yes | no | 248 | – | 248 |
| no | no | yes | 90 | – | 90 |
| no | no | no | 546 | – | 546 |

- Frequencies in 100th generation:

| head definite? | possessor definite? | possessor animate? | prenominal possessor | postnominal possessor | |
|---|---|---|---|---|---|
| | | | | *the* | *a* |
| yes | yes | yes | 1649 | 1829 | – |
| yes | yes | no | 277 | 143 | – |
| yes | no | yes | 46 | 48 | – |
| yes | no | no | 327 | 133 | – |
| no | yes | yes | 0 | – | 238 |
| no | yes | no | 0 | – | 496 |
| no | no | yes | 0 | – | 180 |
| no | no | no | 0 | – | 1092 |

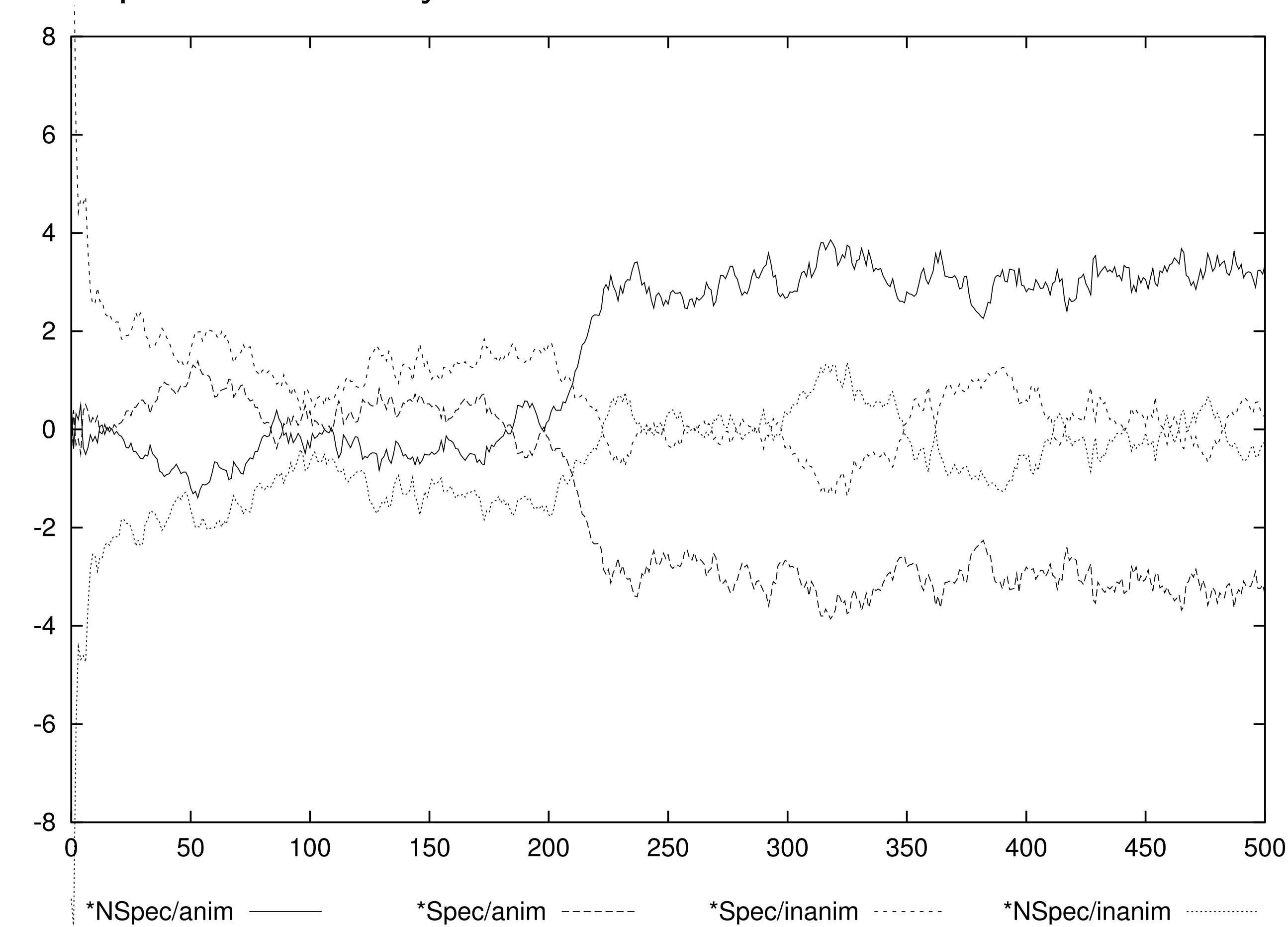
- Frequencies in 200th generation:

| head definite? | possessor definite? | possessor animate? | prenominal possessor | postnominal possessor | |
|---|---|---|---|---|---|
| | | | | *the* | *a* |
| yes | yes | yes | 2604 | 874 | – |
| yes | yes | no | 373 | 47 | – |
| yes | no | yes | 53 | 41 | – |
| yes | no | no | 368 | 92 | – |
| no | yes | yes | 0 | – | 238 |
| no | yes | no | 0 | – | 496 |
| no | no | yes | 0 | – | 180 |
| no | no | no | 2 | – | 1090 |

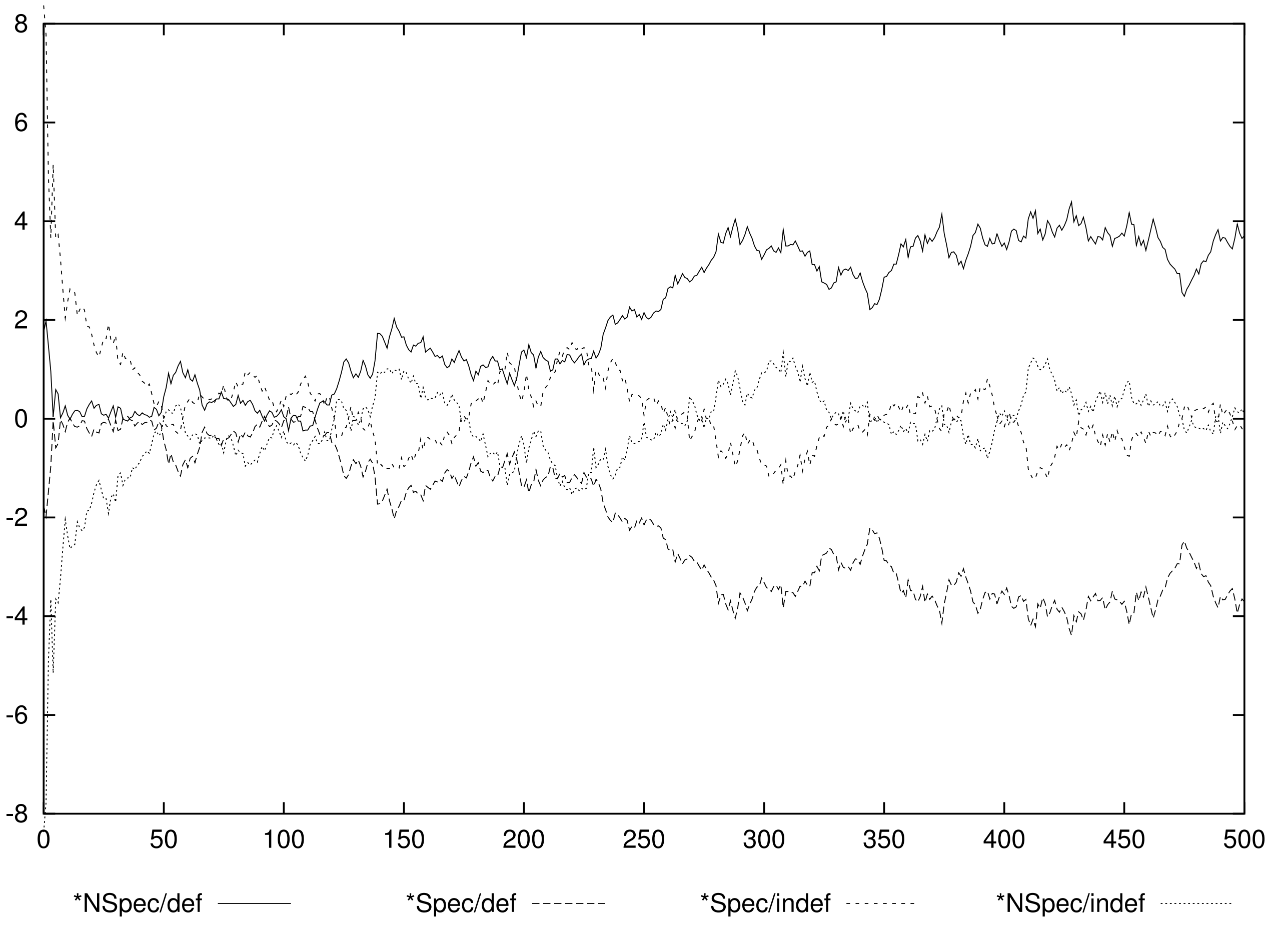
- Frequencies in 300th generation:

| head definite? | possessor definite? | possessor animate? | prenominal possessor | postnominal possessor | |
|---|---|---|---|---|---|
| | | | | *the* | *a* |
| yes | yes | yes | 3448 | 30 | – |
| yes | yes | no | 356 | 64 | – |
| yes | no | yes | 69 | 25 | – |
| yes | no | no | 121 | 339 | – |
| no | yes | yes | 0 | – | 238 |
| no | yes | no | 0 | – | 496 |
| no | no | no | 0 | – | 180 |
| no | no | no | 0 | – | 1092 |

# Development of animacy-related constraints



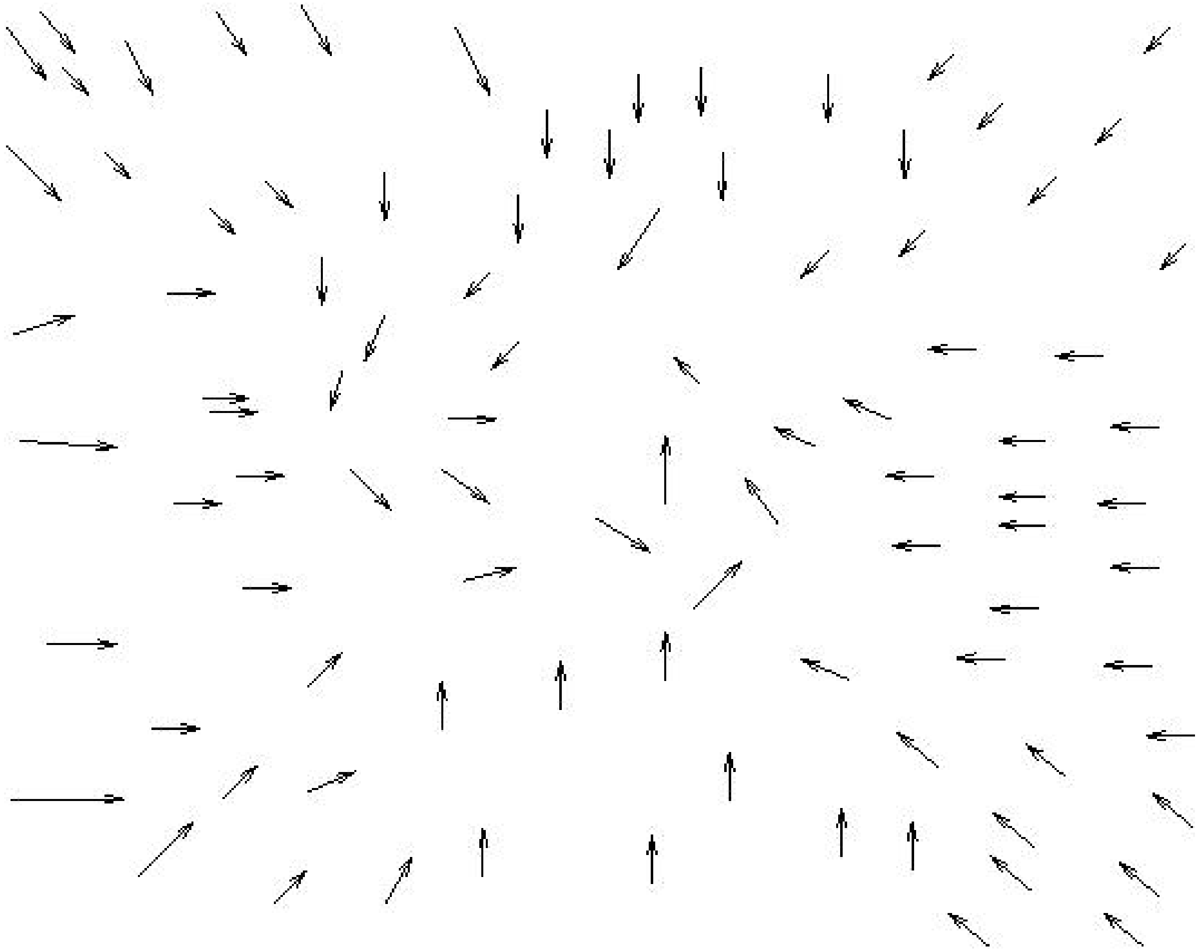
Development of definitess-related constraints



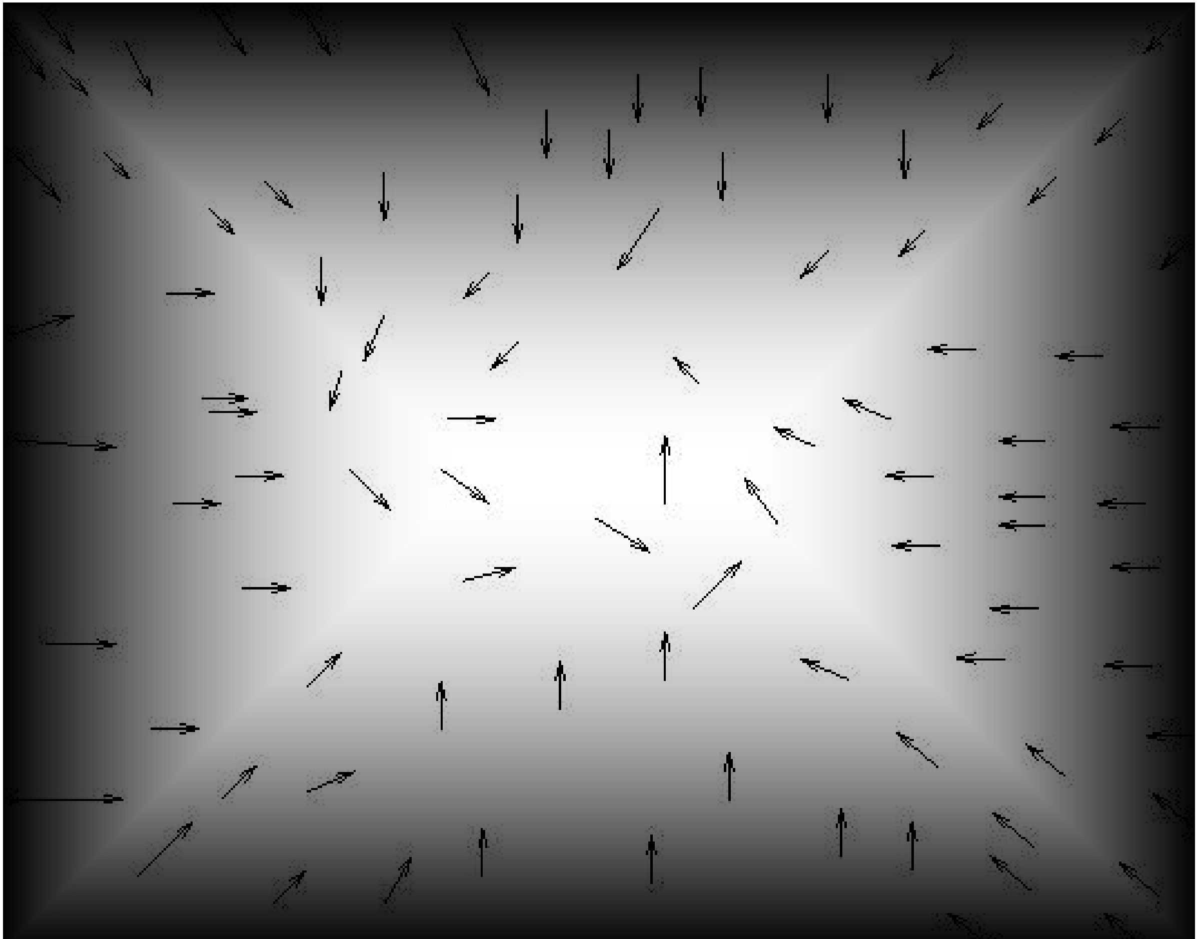
# 6. Conclusion

- possession splits are evolutionary stable
- iterated learning defines dynamics over space of learnable grammars (UG)
- only attainable grammars are expected to occur
- iterated learning makes predictions about which grammars are attainable and which aren't
- predictions about typology and language universals

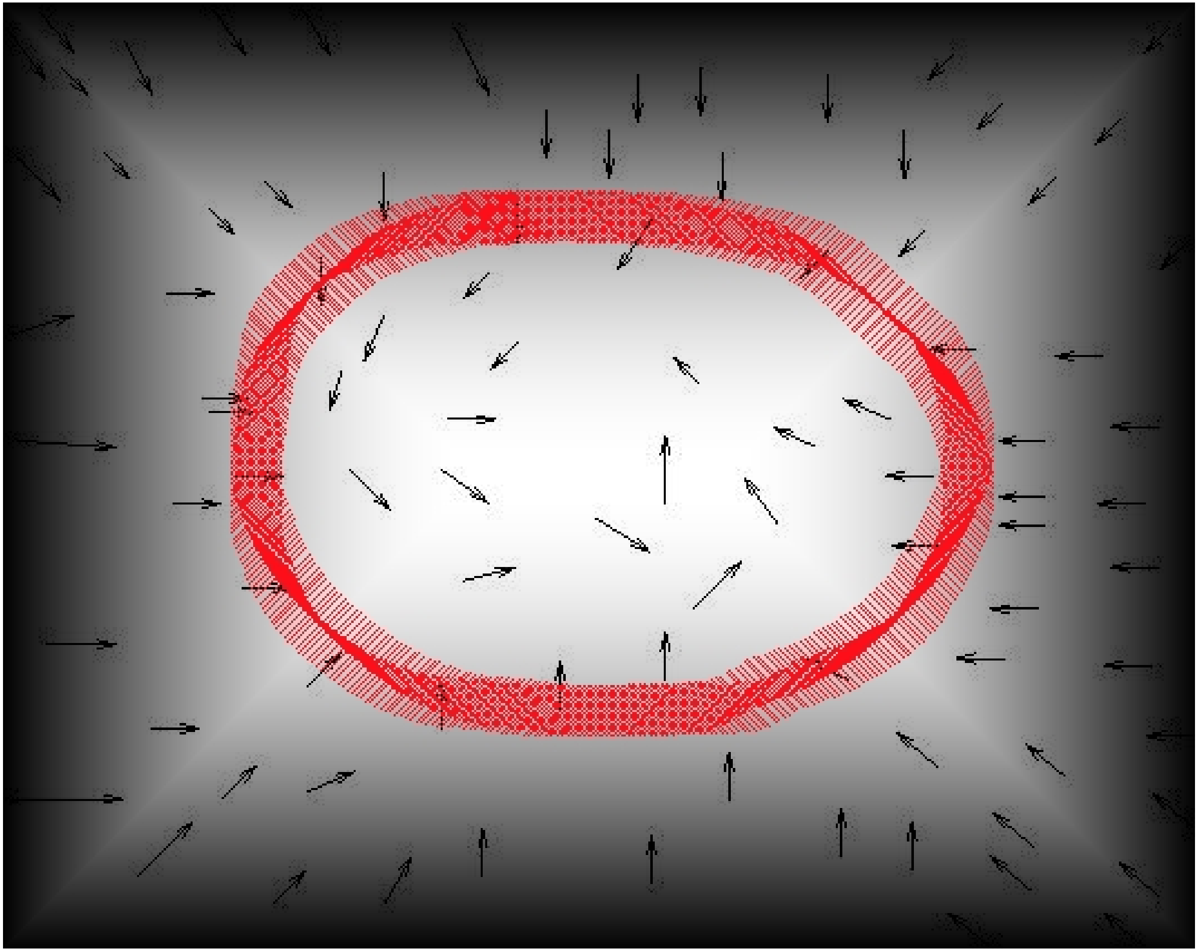
# Universal Grammar



# Universal Grammar



# attainable languages



a possible trajectory