Optimal Communication in a Noisy and Heterogeneous Environment

Willem Zuidema

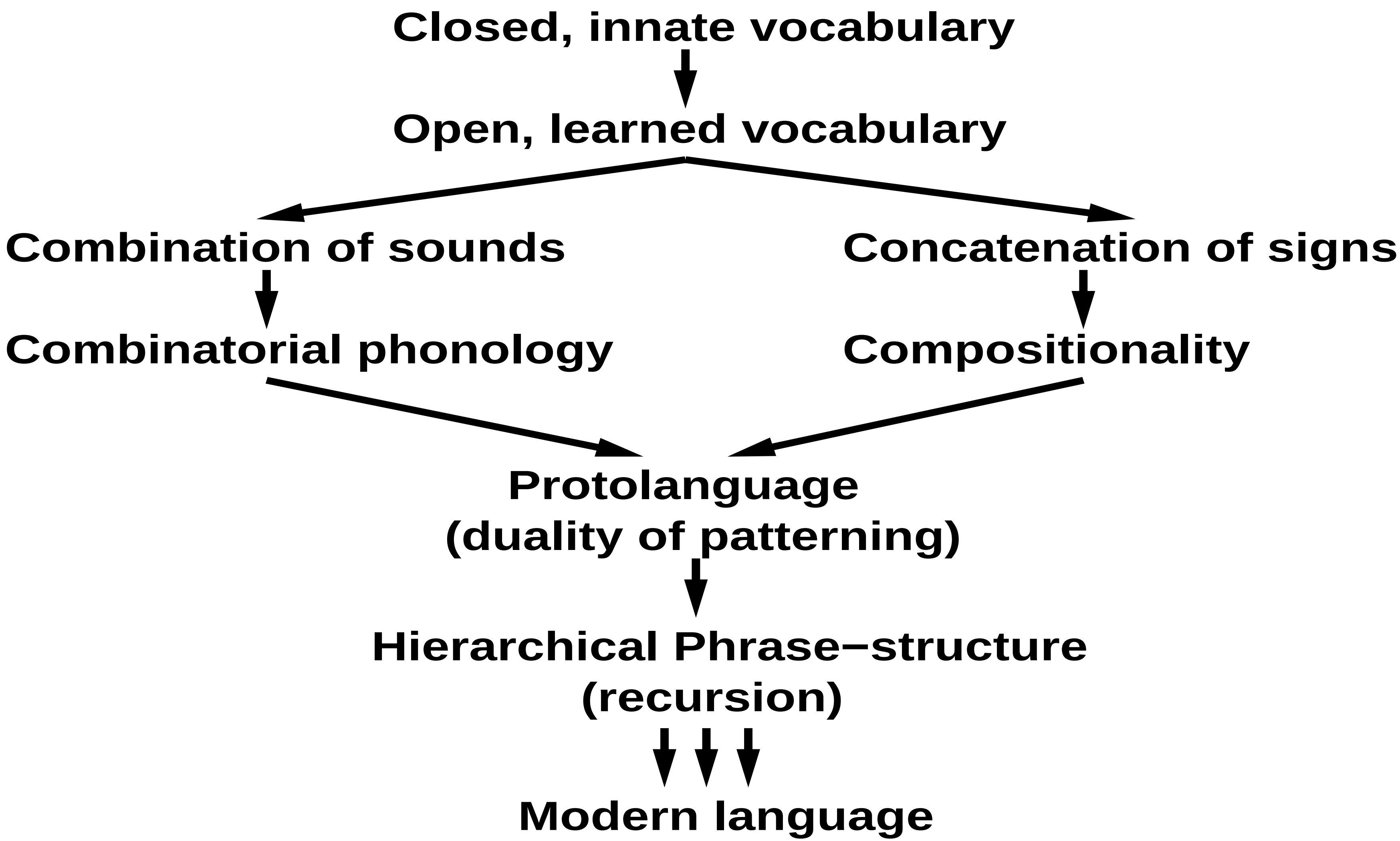
(1) Language Evolution & Computation Research Unit

(2) Institute for Cell, Animal and Population Biology

University of Edinburgh, U.K.

http://www.ling.ed.ac.uk/∼jelle

jelle@ling.ed.ac.uk


Closed, innate vocabulary
Open, learned vocabulary
Combination of sounds
Concatenation of signs
Combinatorial phonology
Compositionality
Protolanguage
(duality of patterning)
Hierarchical Phrase−structure
(recursion)
Modern language

I will not inquire as to the details of how increased expressive power came to spread through a population [...].Accepted practice in evolutionary psychology [...] generally finds it convenient to ignore these problems; I see no need at the moment to hold myself to a higher standard than the rest of the field. (Jackendoff, 2002, Foundations of Language, p. 237)

- Problems of *altruism* and *coordination*

The emergence of compositionality explained?

Compositionality: the property that the meaning of the whole (e.g. a sentence) is a function of the meaning of the parts (e.g. the words) and the way they are put together.

Existing models require a structured language to be already present in the population before the linguistic innovations can successfully spread in a population.

Natural Selection (Nowak & Krakauer, 1999)

Formalism of Hurford (1989, Lingua), Oliphant (1996, PhD-thesis UCSD)

$S =$

intention ↓	sent signal 1*kHz*	2*kHz*	3*kHz*	4*kHz*	5*kHz*
eagle approaches	1.0	0.0	0.0	0.0	0.0
snake approaches	0.0	0.0	1.0	0.0	0.0
tiger approaches	0.0	0.0	0.0	0.0	1.0

$R =$

received signal ↓	interpretation *eagle approaches*	*snake approaches*	*tiger approaches*
1*kHz*	1.0	0.0	0.0
2*kHz*	1.0	0.0	0.0
3*kHz*	0.0	1.0	0.0
4*kHz*	0.0	0.0	1.0
5*kHz*	0.0	0.0	1.0

$$U = \left(\begin{array}{c|ccccc} & \multicolumn{5}{c}{\text{received signal}} \\ \text{sent signal} \downarrow & 1kHz & 2kHz & 3kHz & 4kHz & 5kHz \\ \hline 1kHz & 0.7 & 0.2 & 0.1 & 0.0 & 0.0 \\ 2kHz & 0.2 & 0.6 & 0.2 & 0.0 & 0.0 \\ 3kHz & 0.0 & 0.2 & 0.6 & 0.2 & 0.0 \\ 4kHz & 0.0 & 0.0 & 0.2 & 0.6 & 0.2 \\ 5kHz & 0.0 & 0.0 & 0.1 & 0.2 & 0.7 \end{array}\right)$$

The fitness is given by (Nowak & Krakauer, 1999):

$$F(L, L') = \frac{1}{2} \sum_{m=1}^{M} \sum_{f=1}^{F} \left[S_{mf} \left(\sum_{f'=1}^{F} U'_{ff'} R'_{f'm} \right) + S'_{mf} \left(\sum_{f'=1}^{F} U_{ff'} R_{f'm} \right) \right]$$

Assume the world consists of objects and actions, a fraction ϕ of which is relevant. Nowak & Krakauer (1999) show that with sufficiently high ϕ the maximum fitness of compositional languages is higher. I.e.

$$F(L^+, L^+) > F(L^-, L^-).$$

- However, crucial is that there is a path of *ever increasing fitness* from non-compositional (L^-) to compositional (L^+) languages. I.e.

$$\begin{aligned} F(L^+, L^+) > \quad & F(L^+, L^-) & > F(L^-, L^-) \\ F(L^+, L^+) > \quad & \ldots & > F(L^-, L^-). \end{aligned}$$

Mixed Strategies

$S =$

intention ↓	sent signal A	B	C	ab	cb	ad
1 eagle approaches	$1-x$	0.0	0.0	x	0.0	0.0
2 snake approaches	0.0	$1-x$	0.0	0.0	x	0.0
3 eagle leaves	0.0	0.0	$1-x$	0.0	0.0	x

$R =$

sent ↓	1	2	3
A	1	0	0
B	0	1	0
C	0	0	1
ab	1	0	0
cb	0	1	0
ad	0	0	1

$U =$

sent ↓	received signal A	B	C	ab	cb	ad
A	.	$*$	$*$	0	0	0
B	$*$	.	$*$	0	0	0
C	$*$	$*$	.	0	0	0
ab	0	0	0	$*$	.	.
cb	0	0	0	.	$*$	.
ad	0	0	0	.	.	$*$

N & K show that for mixed strategies, more compositional languages will always do better. I.e., if $x' > x > 0$:

$$F(L', L') > F(L', L) > F(L, L).$$

- However, cost of additional system (memory, confusion) and temporal dimension of holistic signals are completely ignored.
- The capacity for compositional analysis is present before a compositional language is established (as well as in the *Iterated Learning Model*, Kirby, 2000)

Topology preservation

I explore a possible route for a structured language (“superficial compositionality”) to emerge without the capacity for compositionality present in the population.

The structure is *topology preservation* between meaning-space and signal-space, i.e. similar meanings are expressed with similar forms (signals).

Can topology preservation emerge as a side-effect of optimising communication under noisy conditions?

A formalism for communication under noisy conditions

- Assume that there are M different meanings that an individual might want to express, and F different signals (forms) that it can use for this task.

- The communication system of an individual is represented with a *production matrix* S (S gives for every meaning m and every signal f, the probability that the individual chooses f to convey m);

- and an *interpretation matrix* R. (R gives for every signal f and meaning m, the probability that f will be interpreted as m).

- Signals can be more or less similar to each other and there is noise on the transmission of signals which depends on these similarities (*confusion matrix* U).

- Meanings can be more or less similar to each other, and the value of a certain *interpretation* depends on how close it is to the *intention* (*value matrix* V)

Example: Vervet monkey alarm calls

Three different types of predators: from the air (eagles), from the ground (leopards) and from the trees (snakes).

The monkeys are capable of making a number (say 5) of different sounds that range on one axis (e.g. pitch, from high to low) and are more easily confused if they are closer together.

If one makes a mistake, typically not every mistake is equally bad.

$U =$

sent signal ↓	received signal $1kHz$	$2kHz$	$3kHz$	$4kHz$	$5kHz$
$1kHz$	0.7	0.2	0.1	0.0	0.0
$2kHz$	0.2	0.6	0.2	0.0	0.0
$3kHz$	0.0	0.2	0.6	0.2	0.0
$4kHz$	0.0	0.0	0.2	0.6	0.2
$5kHz$	0.0	0.0	0.1	0.2	0.7

$V =$

interpretations ↓	intentions *eagle*	*snake*	*leopard*
eagle	0.9	0.5	0.1
snake	0.2	0.9	0.2
leopard	0.1	0.5	0.9

$$S = \left(\begin{array}{c|ccccc} & \multicolumn{5}{c}{\text{sent signal}} \\ \text{intention} \ \downarrow & 1\textit{kHz} & 2\textit{kHz} & 3\textit{kHz} & 4\textit{kHz} & 5\textit{kHz} \\ \hline \textit{eagle} & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ \textit{snake} & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ \textit{leopard} & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{array} \right)$$

$$R = \left(\begin{array}{c|ccc} & \multicolumn{3}{c}{\text{interpretation}} \\ \text{received signal} \ \downarrow & \textit{eagle} & \textit{snake} & \textit{leopard} \\ \hline 1\textit{kHz} & 1.0 & 0.0 & 0.0 \\ 2\textit{kHz} & 1.0 & 0.0 & 0.0 \\ 3\textit{kHz} & 0.0 & 1.0 & 0.0 \\ 4\textit{kHz} & 0.0 & 0.0 & 1.0 \\ 5\textit{kHz} & 0.0 & 0.0 & 1.0 \end{array} \right)$$

$$F_{ij} = V \cdot \left(S^i \times \left(U \times R^j\right)\right) \tag{1}$$

In this formula, "$\times$" represents the usual matrix multiplication and "$\cdot$" represents dot-multiplication (the sum of all multiplications of corresponding elements in both matrices; the result of dot-multiplication is not a matrix, but a scalar).

$$\begin{aligned} F_{ij} = 0.7 \quad & \times \quad 0.9 + 0.2 \times 0.5 + 0.2 \times 0.5 + 0.6 \times 0.9 \\ & + \quad 0.2 \times 0.5 + 0.1 \times 0.5 + 0.2 \times 0.9 + 0.7 \times 0.9 = 2.33 \end{aligned}$$

Distributed hill-climbing

- The values in the S and R matrices are all either 1 or 0
- Distributed hill-climbing:
  1. Random speaker (i) and hearer (j) are picked, and F_{ij} is measured;
  2. A random change is made in a random matrix of the speaker (or hearer), and F_{ij} is measured again;
  3. If the F_{ij} is better, the change is kept; otherwise, it is reverted.

Motivation

for this style of optimization:

1. it is fast and straightforward to implement;

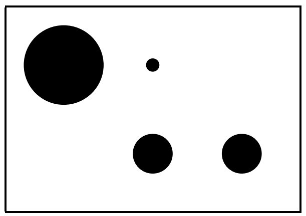
2. it works well, and gives, if not the optimum, a good insight on characteristics of the optimal communication system;

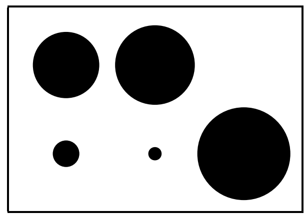
3. it shows possible *routes* to (near-) optimal communication systems, and in a sense forms an abstraction for both learning and evolution.

Visualising the results

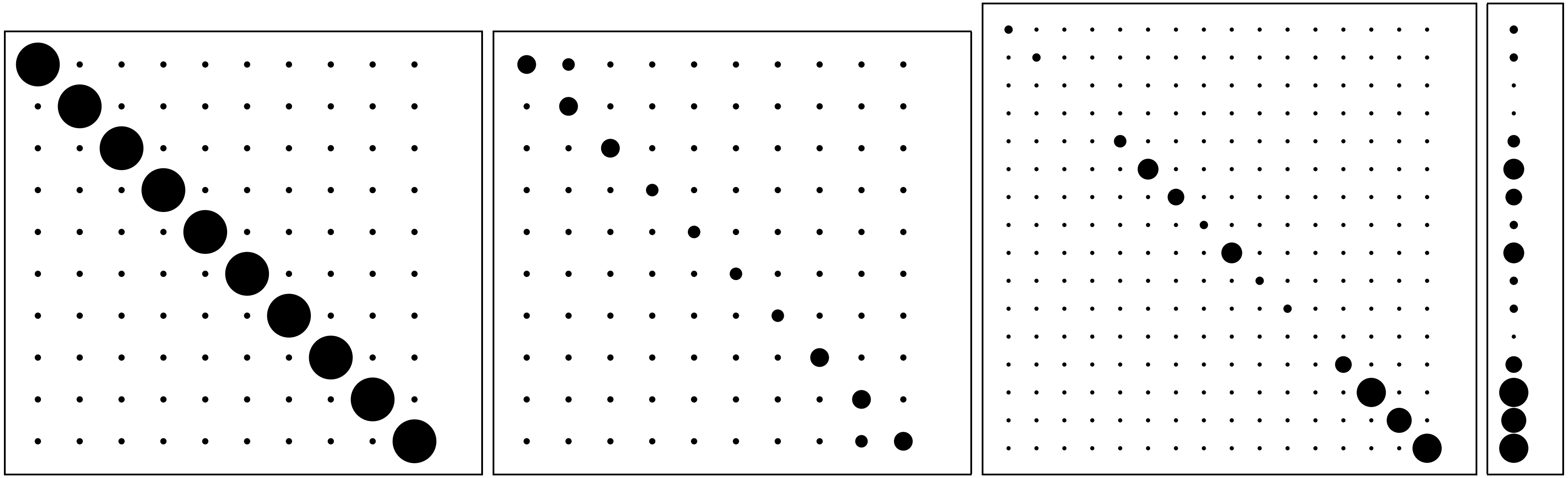
$$S = \left(\begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline m_1 & 0.9 & 0.1 & 0.0 \\ m_2 & 0.0 & 0.5 & 0.5 \end{array} \right) \qquad R = \left(\begin{array}{c|cc} & m_1 & m_2 \\ \hline f_1 & 0.7 & 0.3 \\ f_2 & 0.9 & 0.1 \\ f_3 & 0.0 & 1.0 \end{array} \right)$$

$$R^T = \left(\begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline m_1 & 0.7 & 0.9 & 0.0 \\ m_2 & 0.3 & 0.1 & 1.0 \end{array} \right)$$



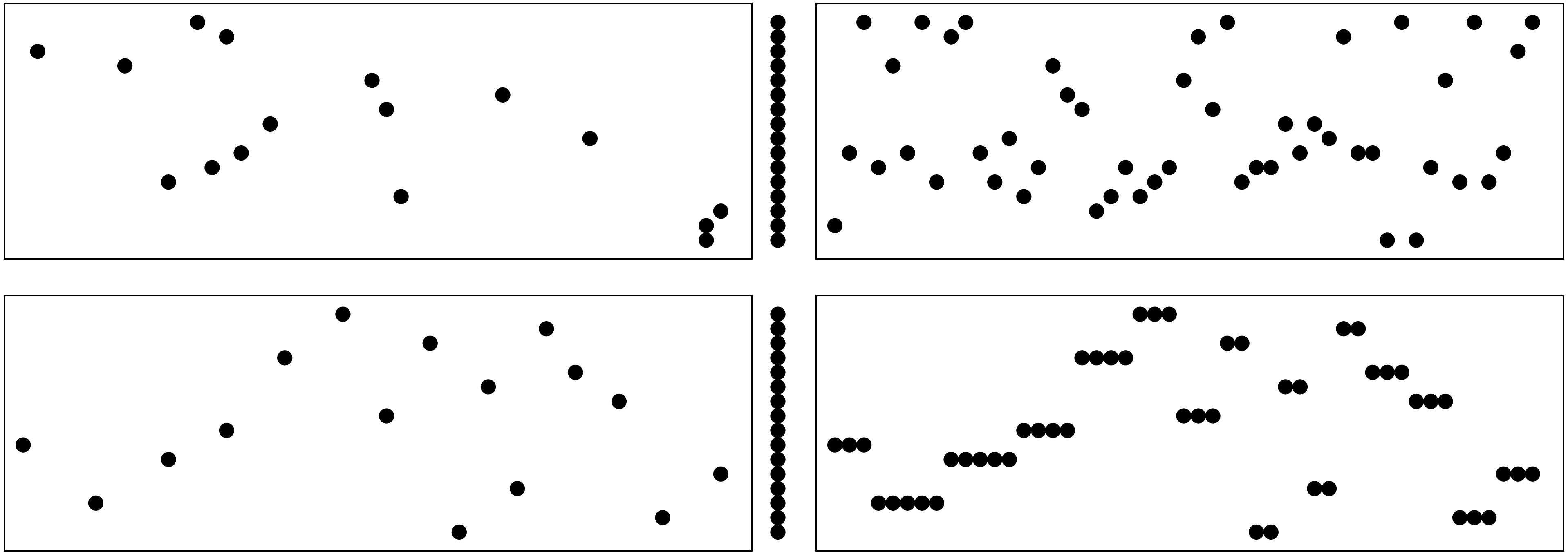


Control parameters: V and U-matrices



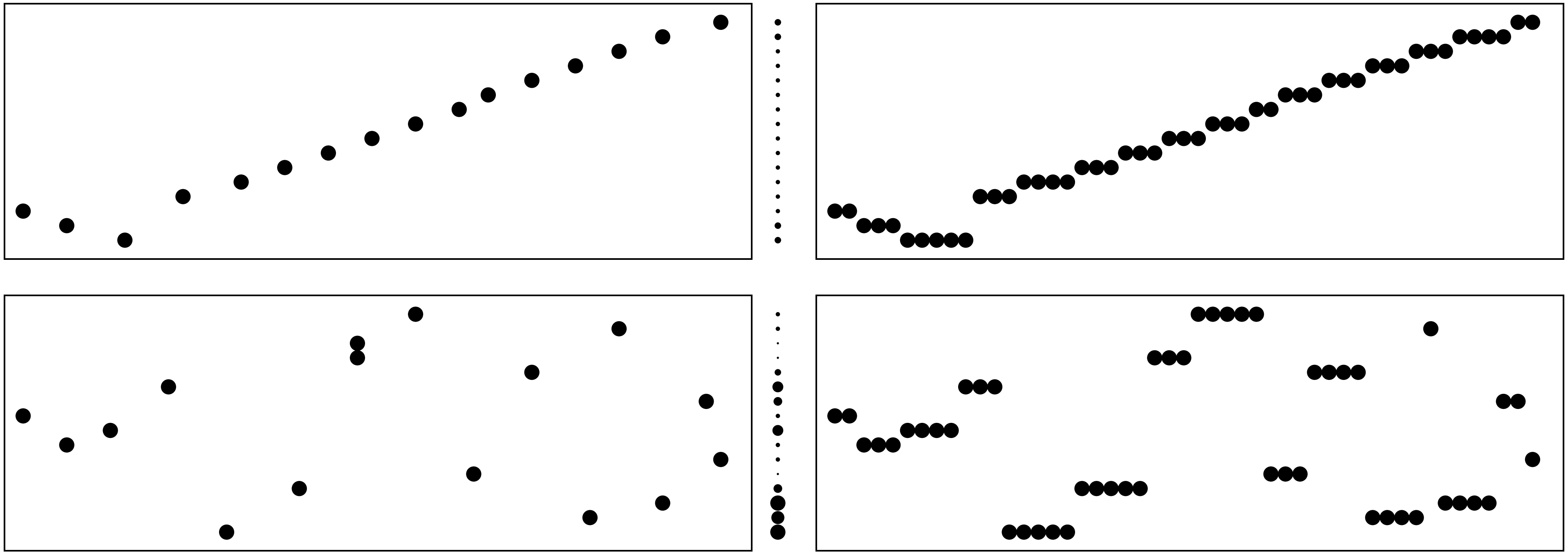
V:0d, 1d homogeneous, 0d heterogeneous

Specificity, Coherence, Distinctiveness



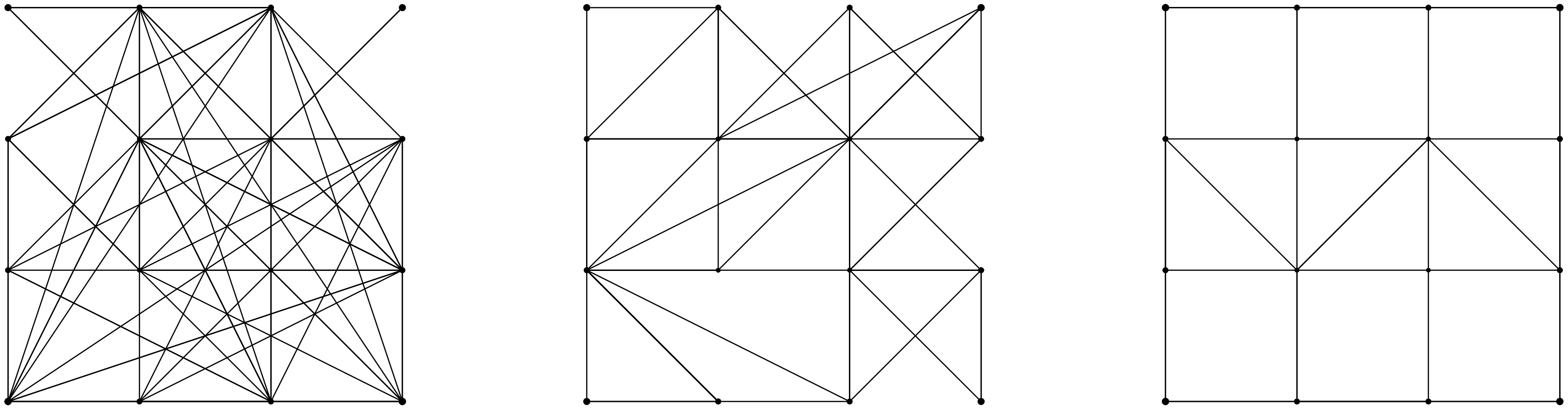
U:1d, V:0d homogeneous, $t = 0$, $t = \infty = 2 \times 10^8$

Topology preservation, Heterogeneity



U:1d, V:0d homogeneous/heterogeneous, $t = \infty$

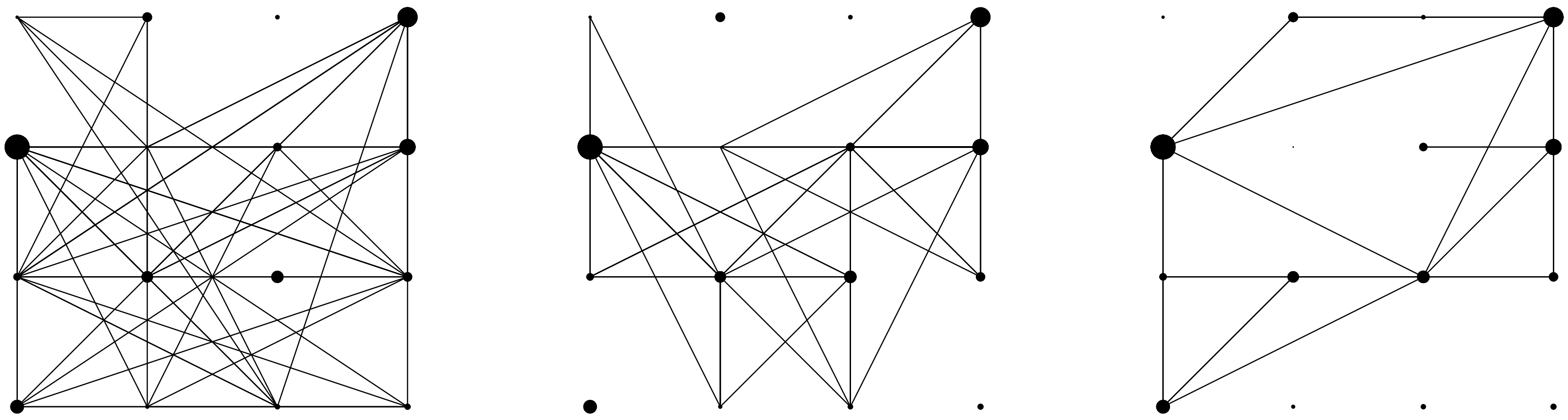
Results: 2d meaning spaces



(a) t=0 (b) t=10^6 (c) t=∞

U:2d, V:2d homogeneous, $t = 0, 10^6, \infty$
(Points in meaning space are connected, if their preferred forms are neighbours in form space)

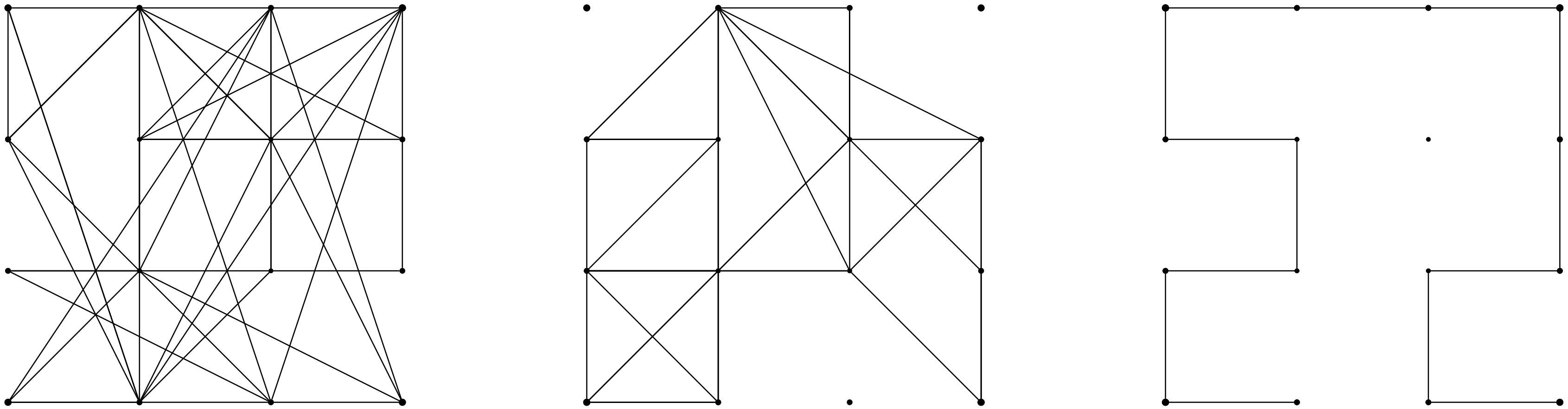
Heterogeneity: sacrificing low-valued meanings



(a) t=0 (b) t=10^6 (c) t=∞

U:2d, V:2d heterogeneous, $t = 0, 10^6, \infty$

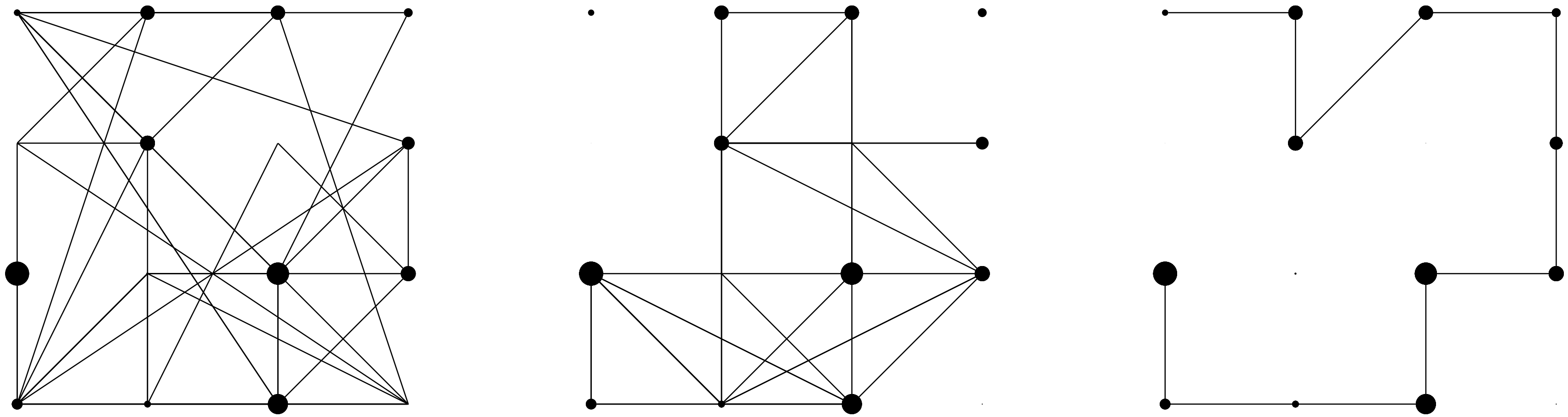
Dimensionality mismatch



(a) t=0 (b) t=10^6 (c) t=∞

U:1d, V:2d homogeneous, $t = 0, 10^6, \infty$

Dimensionality mismatch & Heterogeneity



(a) t=0 (b) t=10^6 (c) t=∞

U:1d, V:2d heterogeneous, $t = 0, 10^6, \infty$

Summary of results

- Specificity
- Coherence (Lewis, 1969; Steels, 1996; Oliphant, 1996)
- Distinctiveness (De Boer, 1999; Nowak & Krakakauer, 1999)
- Topology preservation (Zuidema & Westermann, 2003)
- High-value meanings first
- Low-value meanings sacrificed

Discussion

- Evolutionary Game Theory: *Necessary and sufficient conditions for evolutionary stable languages ($\sim$ Matina Donaldson, p.c.), dynamic analysis*;

- Information Theory: *Maximum fitness of languages in a noisy (U) and heterogeneous (V) environment ($\sim$ Plotkin & Nowak, 2000)*;

- Evolution of Language: *does topology preservation facilitate the spread of the capacity for compositionality?*

- Sound symbolism?

Conclusions

1. Crucial for evolutionary explanations of all aspects of language, is to explain how linguistic innovations can spread in a population; showing a better end result is neither sufficient nor necessary;

2. Including plausible assumptions on noise in signalling and a topology in the meaning space, as studied in a simple simulation, opens up the possibility for rich pattern formation that was overlooked in previous robotic and mathematical models;

3. Combinatorial patterning as a strategy to minimize the effects of noise is a possible precursor for *productive* combination;

4. A rich formalism allows for side-effects in evolutionary optimization; side-effects of one adaptation (e.g. learning, noise robustness) might facilitate the next (e.g. phonemic coding, compositionality).

Acknowledgments & References

- supervisors: **Simon Kirby**, **Jim Hurford** (LEC, Edinburgh), **Nick Barton** (ICAPB, Edinburgh)

- with **Gert Westermann**, “Evolution of an Optimal Lexicon under Constraints from Embodiment”, *Artificial Life*, to appear
- with **Bart de Boer**, “How did we get from there to here in the evolution of language?”, *Behavioral and Brain Sciences*, to appear
- with **Nick Barton** (2003), “Evolution: the erratic path towards complexity”, *Current Biology*, 13:16
- Ray Jackendoff (2002), “Foundations of Language”, Oxford Uni. Press
- Martin Nowak & David Krakauer (1999), “The evolution of language”, *Proc. Nat. Acad. Sci. USA* 96
- Simon Kirby (2002), “Natural Language from Artificial Life”, *Artificial Life* 8:2.

- Funding: University of Edinburgh, E.U. Marie Curie fellowship, Prins Bernhard Culture Fund Amsterdam.