EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Language Evolution WiSe 2023/2024

Lecture 11: Experiments

28/11/2023, Christian Bentz

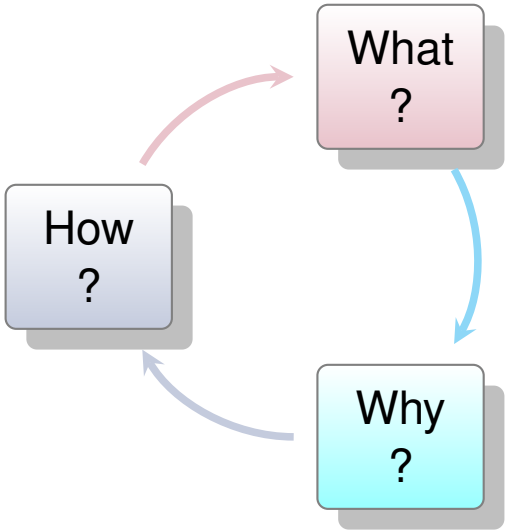
Overview


photo by Abdullah Mohiuddin
CC BY-NC-SA Hajime Yamauchi
GGCCTGACACCTTCGCACC

Recap

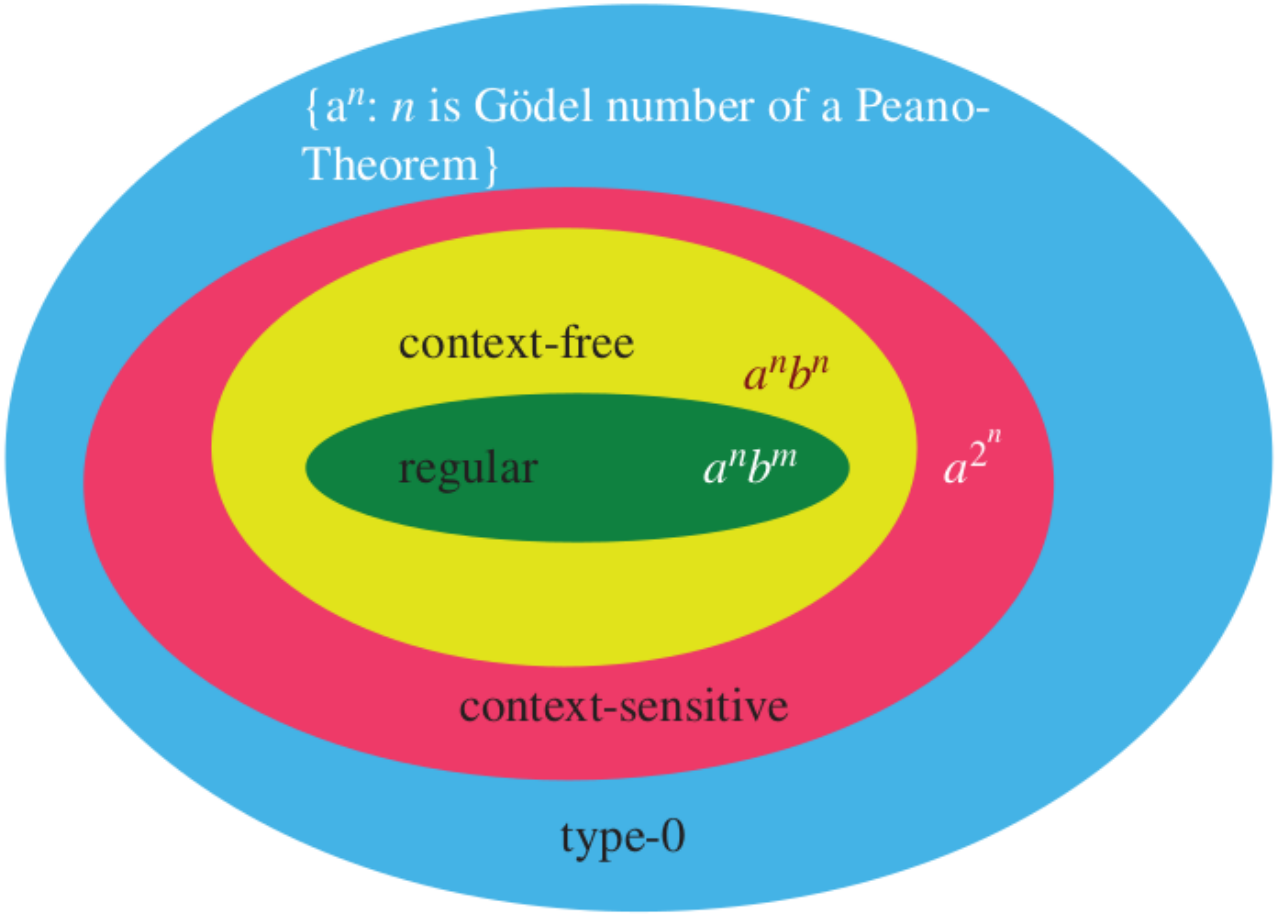
Three Questions

1. **What** evolved, i.e. what is “language” in the first place?
2. **Why** did it evolve, i.e. did it have particular functions?
3. **How** did it evolve?



What is Language?

The Classical Chomsky Hierarchy



Jäger & Rogers (2012), p. 1959.

Recursion (Important Take-Home-Message)

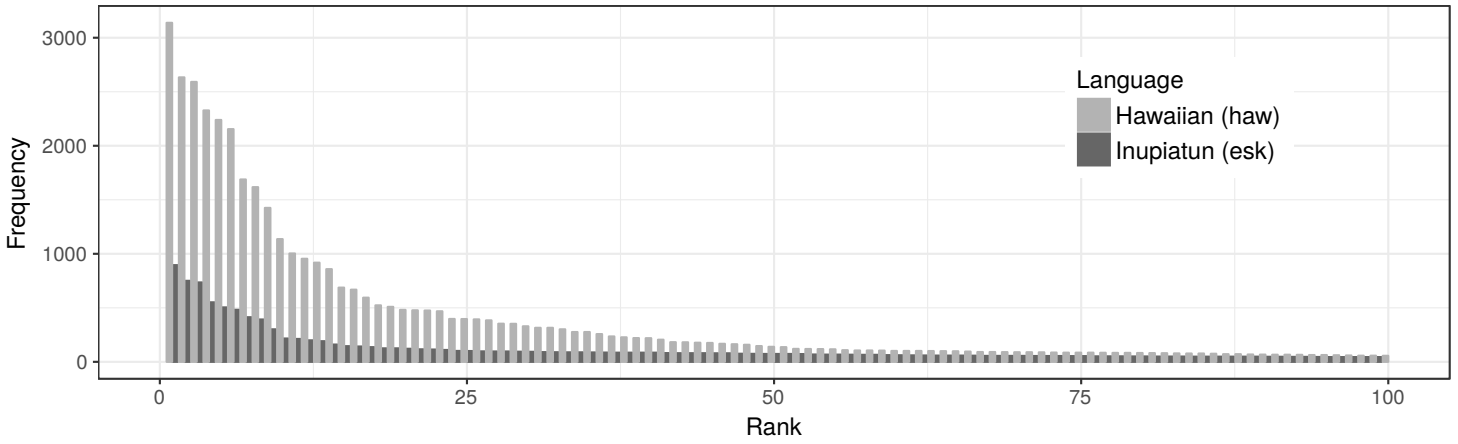
We can introduce **recursion** into a formal grammar by any rule which has the same non-terminal(s) on the left and right hand side:

$$\square \rightarrow \triangleleft \square \tag{1}$$

So this is already possible in **regular** grammars, i.e. the lowest level of the traditional Chomsky Hierarchy. Arguably, there are some natural language structures where such a recursive pattern is needed. For instance, when a number of adjectives (potentially arbitrarily large) is added before a noun in the English noun phrase (e.g. *a bright, friendly, welcoming, ... friend*).

$$\overline{N} \rightarrow A\,\overline{N} \tag{2}$$

Word Frequency Distributions



Hawaiian (haw)

40001001 O ke kuauhau na ka hanauna o Iesu Kristo , ka mamo a Davida , ka mamo a Aberahama.

40001002 Na Aberahama o Isaaka ; na Isaaka o Iakoba ; na Iakoba o Iuda a me kona poe hoahanau;

[...]

Iñupiatun (esk)

40001001 Uvva ukua aglang ich sivulliang iñ Jesus Christ-ng um , kinguviang upluni David-miñ Abraham-miñḷu .

40001002 Abraham aapagigaa Isaac-ng um , Isaac-li aapagigaa Jacob-ng um , Jacob-li aapagigaa Judah-ng um aniqataiñḷu .

Mayer and Cysouw (2014). A massively parallel Bible corpus.

Zipf's Law (of Word Frequencies)

Word	Rank	Freq	Char
the	1	12539	3
and	2	9964	3
of	3	7459	2
to	4	7317	2
in	5	3985	2
you	6	3747	3
for	7	3014	3
is	8	2957	2
he	9	2925	2
a	10	2862	1
...	...	...	...
work–then	2742	1	10
world-rulers	2743	1	12
worm	2744	1	4
wormwood	2745	1	8
wounding	2746	1	8
writer	2747	1	6
writers	2748	1	7
zarephath	2749	1	9
zenas	2750	1	5

Another (more common) formulation of the law:

$$f(w) \propto \frac{1}{r^{\alpha}} \tag{3}$$

The α-paramter is the slope in log-log space (i.e. when both the ranks and frequencies are log transformed). Zipf assumed that $\alpha \sim 1$ holds across languages.

Unigram Entropy

The unigram entropy is the **average information content** of all types.

$$H(X) = -\sum_{i=1}^{V} p(x_i) \log_2 p(x_i),$$

- X: random variable drawn from the set of types (i.e. $\mathcal{V}$),
- V: number of types (as before).

Shannon, Claude E. (1948). A mathematical theory of communication.
Cover & Thomas (2006). Elements of information theory, p. 14.

Example (Characters)

```
All human beings are born free and
equal in dignity and rights
```

unit	char.freq
a	5
A	1
b	2
d	3
e	5
f	1
...	...

$\widehat{H}(X) = -(\frac{5}{51} \log_2(\frac{5}{51}) + \frac{1}{51} \log_2(\frac{1}{51}) + \dots) \sim$ 3.97 bits/char

Note: This example uses the so-called *maximum likelihood* (ML) estimator for probabilities. This gives the estimated $\hat{p}$ and $\widehat{H}$.

Example: Bird Song and Human Language

```
rn rn kd rq rp km jx km rn rn kd rq rp ro as rr rs rt
ls as am rn rn kd rq rp ro ro lo rn rn kd rq rp as rr
   rs rt rh rn rn tw nn ir rh tx rn lo rs rt rh
```

$$\widehat{H}(X) \sim 3.1 \text{ bits/char}$$
$$\widehat{H}(X) \sim 3.9 \text{ bits/char.string}$$

```
All human beings are born free and equal in dignity
and rights.  They are endowed with reason and
conscience and should act towards one another in a
spirit of brotherhood
```

$$\widehat{H}(X) \sim 4.1 \text{ bits/char}$$
$$\widehat{H}(X) \sim 4.5 \text{ bits/char.string}$$

How and *why* did this *uniqueness* evolve?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

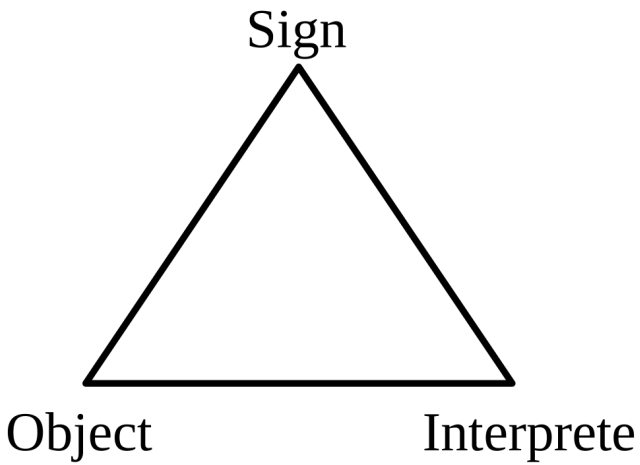
Introduction

Semiotics (Peirce)

"A **sign** is something which is in a **triadic relation** to two things being a **sign** of an **object** for an **interprete**."

Peirce (2016). Prolegomena to a science of reasoning, p. 62.

Note: This is different from the **dyadic** relationship between *signifié* and *signifiant* promoted by Ferdinand De Saussure.



Semiotics (Peirce)

"I extend logic to embrace all the necessary principles of semeiotic, and I recognize a logic of **icons**, and a logic of **indices** as well as a logic of **symbols**; [...]"

Peirce (2016). Prolegomena to a science of reasoning, p. 86.

Note: In Peirce's terminology these are all *signs*.

huǒ

火

Fire

Icon: A sign which represents the object by means of **resemblance/similarity**.

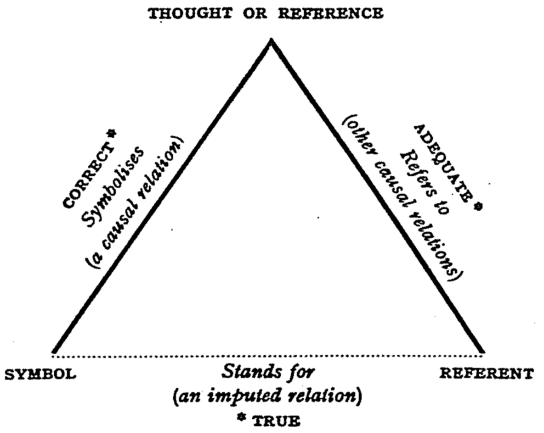
Index: A sign which is typically in a **co-occurrence relation** with the object it represents.

Symbol: A sign with an **arbitrary relation** to an object, **conventionally** used by interpretes to be also understood by other interpretes.

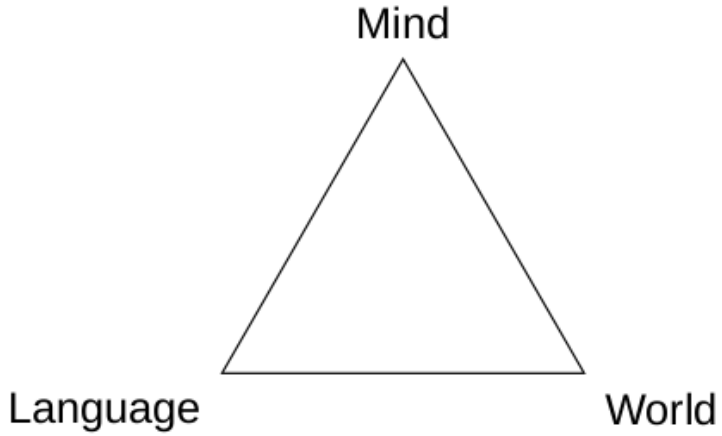
Semiotics (Human Language)

"**Semiotics** is the study of the relationship between **signs and their meanings**. In this book we are interested in the relationship between forms and meanings in certain kinds of symbolic systems, namely human languages. The diagram is a way of illustrating how speakers use language to describe things, events, and situations in the world."

Kroeger (2019). Analyzing meaning, p. 16.



Ogden & Richards (1923).
The meaning of meaning, p. 11.



Kroeger (2019). Analyzing meaning, p. 16.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

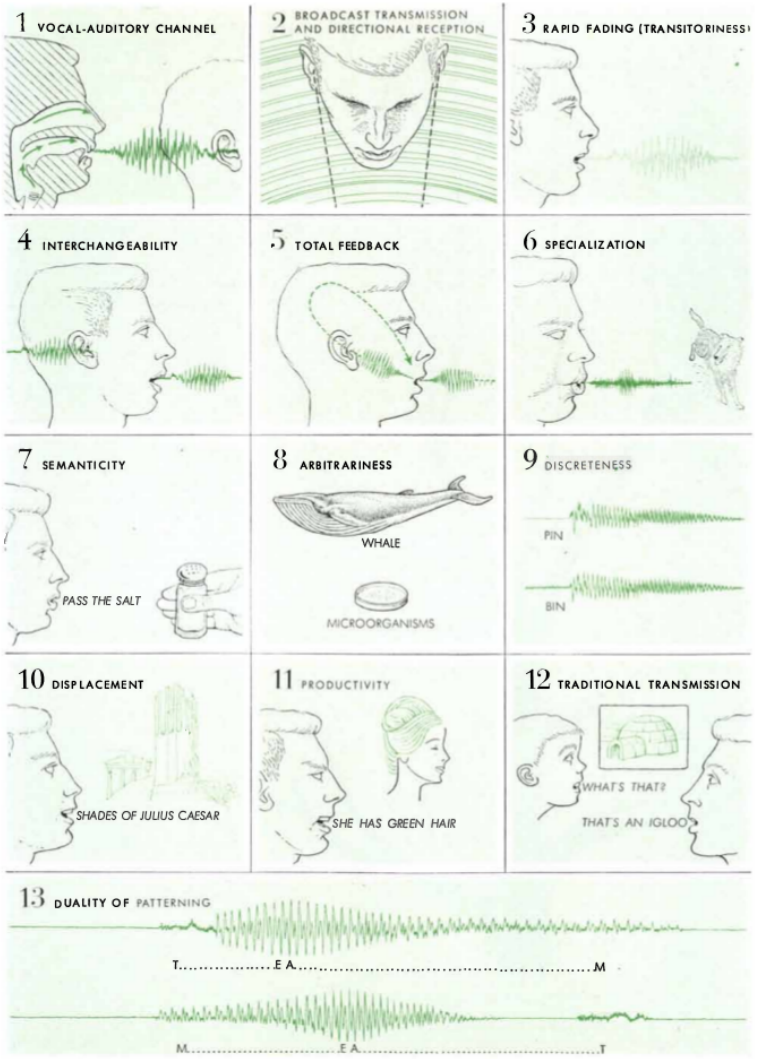
Faculty of Philosophy
General Linguistics

Combinatoriality and Compositionality

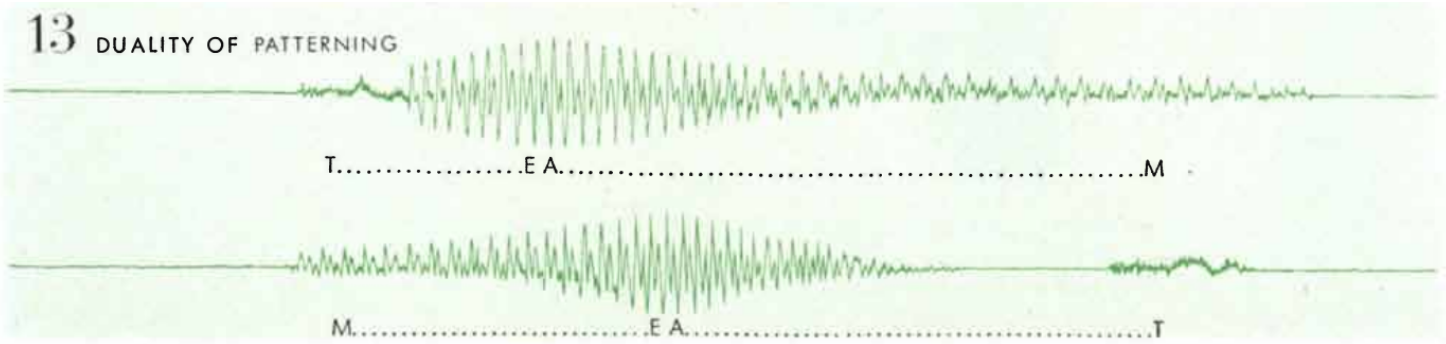
The Design Features of Human Language

"A set of 13 design-features is presented in the illustration on the opposite page. There is solid empirical justification for the belief that all the languages of the world share every one of them."

Hockett (1960). The origin of speech, p. 90.



Duality of Patterning



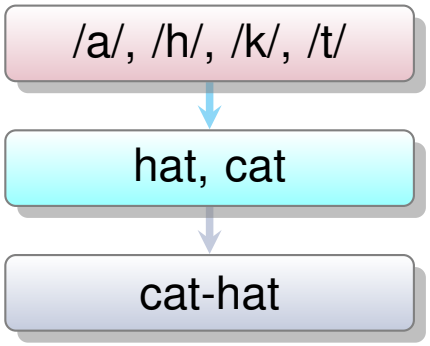
"The **meaningful** elements in any language [...] constitute an enormous stock. Yet they are represented by small arrangements of a relatively very small stock of distinguishable sounds which are themselves wholly **meaningless**."

Hockett (1960). The origin of speech, p. 90.

Duality of Patterning

"Language is structured on at least two levels (Hockett, 1960). On one level, a small number of meaningless building blocks (phonemes, or parts of syllables for instance) are combined into an unlimited set of utterances (words and morphemes). This is known as **combinatorial structure**. On the other level, meaningful building blocks (words and morphemes) are combined into larger meaningful utterances (phrases and sentences). This is known as **compositional structure**."

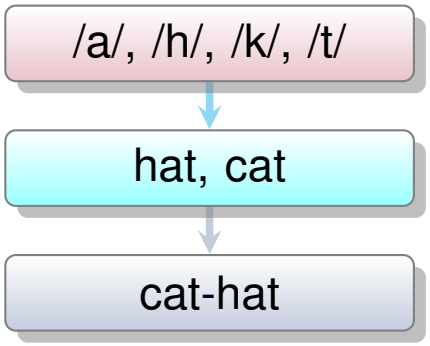
Little et al. (2017), p. 1.



Duality of Patterning

"Language is structured on at least two levels (Hockett, 1960). On one level, a small number of meaningless building blocks (phonemes, or parts of syllables for instance) are combined into an unlimited set of utterances (words and morphemes). This is known as **combinatorial structure**. On the other level, meaningful building blocks (words and morphemes) are combined into larger meaningful utterances (phrases and sentences). This is known as **compositional structure**."

Little et al. (2017), p. 1.



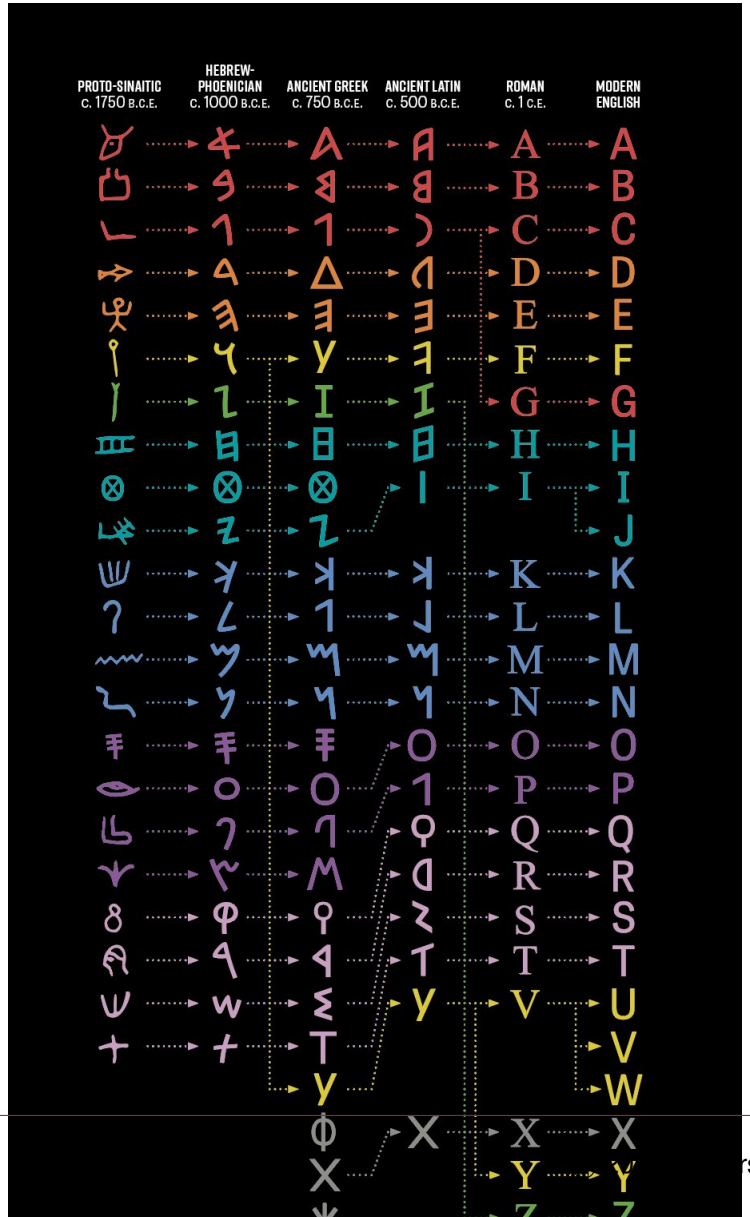
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 1: Experimental Semiotics

Icon → Symbol

It is often assumed that *graphical*, *auditory*, and *gestural* signals evolved from **icons**, i.e. depictions of the objects/concepts denoted. Over time, as they were used more often, they became more simplified and abstract, and finally **symbols** with an arbitrary relationship to the objects they signify.



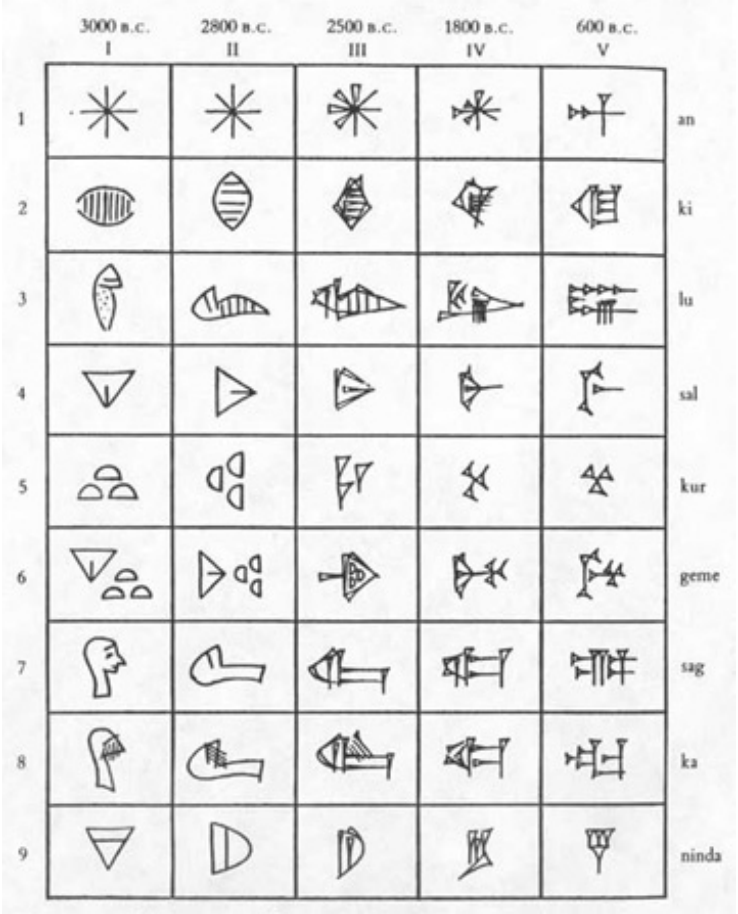
Icon → Symbol

It is often assumed that *graphical*, *auditory*, and *gestural* signals evolved from **icons**, i.e. depictions of the objects/concepts denoted. Over time, as they were used more often, they became more simplified and abstract, and finally **symbols** with an arbitrary relationship to the objects they signify.

https://www.omniglot.com/chinese/evolution.htm

Icon → Symbol

It is often assumed that *graphical*, *auditory*, and *gestural* signals evolved from **icons**, i.e. depictions of the objects/concepts denoted. Over time, as they were used more often, they became more simplified and abstract, and finally **symbols** with an arbitrary relationship to the objects they signify.



Research Question

Can we reconstruct how this transition from icons to symbols happens?

Language Evolution in the Lab

Over the last c. 15 years a paradigm has emerged for testing language evolution models in the lab: **experimental semiotics**. Various experimental designs can be used in this context.

Nölle and Galantucci (2022). Experimental semiotics: past, present, and future.

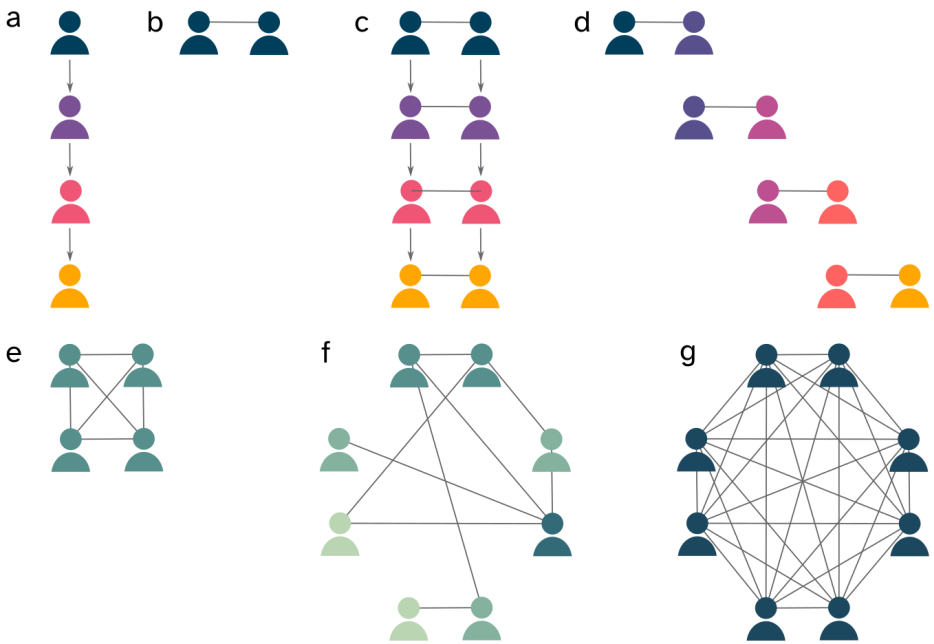


Figure: Experimental designs. Arrows indicate vertical transmission; solid lines indicate horizontal interaction. a) Transmission in an iterated learning chain. b) Pairs of participants (so-called dyads) solve a communication game. c) Transmission and interaction: Dyads interact within an iterated chain, where the output of a dyad becomes the input for the next generation. d) Replacement method: Dyads interact, and after some time one dyad member is replaced with a new participant. e-f) Micro-societies: Participants interact with several members of an artificial 'community' to simulate the spread of variants in a population. Different population sizes (e.g., e vs f) and social network topologies (e.g., fully connected networks in e and g vs a sparse network in f) allow studying the effect of population dynamics on communication systems.

Evolution of Symbols in the Lab

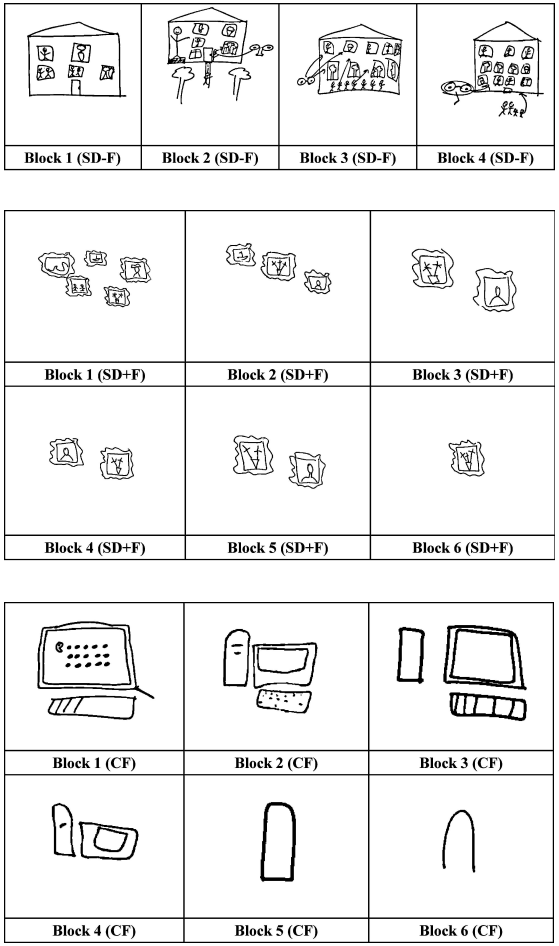
"Like the game *Pictionary*, participants were required to depict various concepts in such a way that a partner could identify them. Again, like Pictionary, participants were not allowed to speak or use letters in their drawings. Unlike Pictionary, concepts were drawn from a list of 16 items, which were known to both participants. Concepts included easily confusable items such as drama, soap opera, theatre [...]"

Garrod et al. (2007). Foundations of representation: Where might graphical symbol systems come from?

Places	People	Programmes	Objects	Abstract
Theatre	Robert De Niro	Drama	Television	Loud
Art gallery	Arnold Schwarzenegger	Soap opera	Computer monitor	Homesick
Museum	Clint Eastwood	Cartoon	Microwave	Poverty
Parliament				

Results

- In conditions *without interaction* (SD-F) the drawings became more **complex** and retained their **iconic** character.
- In conditions with feedback (SD+F) and interaction (CD+F, DD+F) the drawings became **simpler** and more **symbolic**.

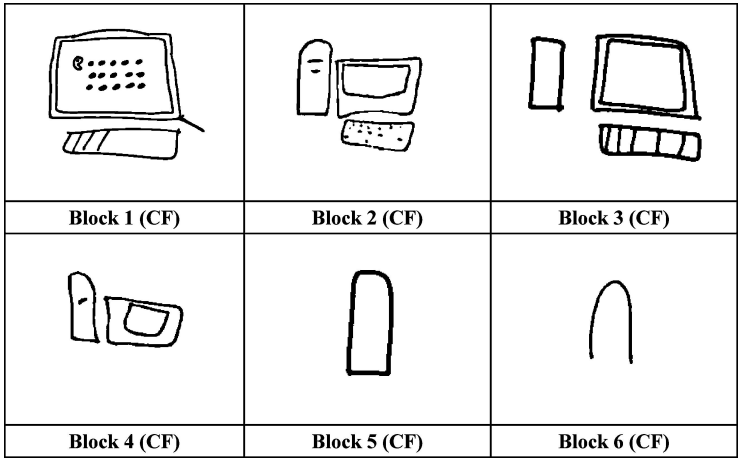


Upper panel: Experiment with single director (SD), i.e. single drawer, and no feedback (-F). Middle panel: Single director (SD) with feedback (+F). Lower panel: Concurrent feedback (CF).

Sign ↔ Mind

"On the basis of these results, this article argues that icons evolve into symbols as a consequence of the systematic shift in the **locus of information** from the **sign** to the user's memory of the sign's usage supported by an interactive grounding process."

Garrod (2007), p. 961.



Further Studies

Gestural communication

Little et al. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*.

Auditory communication

Verhoef et al. (2015). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science*.

Graphic communication

Morin (2018). Spontaneous emergence of legibility in writing systems: The case of orientation anisotropy.

Roberts et al. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*.

Galantucci et al. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies*.

Galantucci et al. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*.

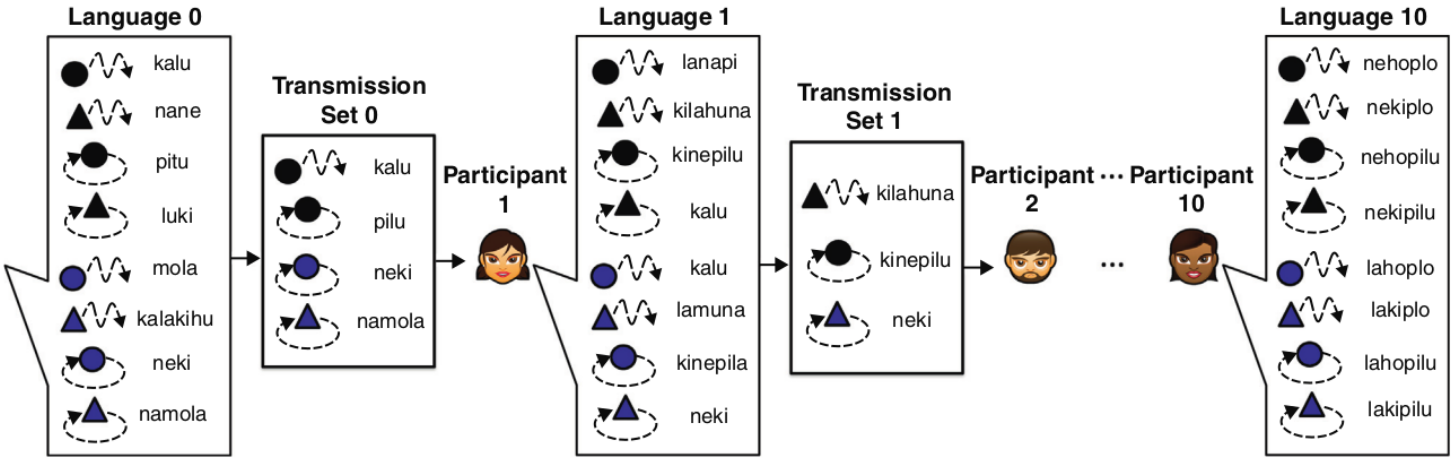
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 2: Iterated Learning

Iterated Learning: Definition

"The process by which a behaviour arises in one individual through induction on the basis of observations of behaviour in another individual who acquired that behaviour in the same way."



Note: *Language 0* is typically a random mapping of randomly generated strings to shape-color-movement combinations.

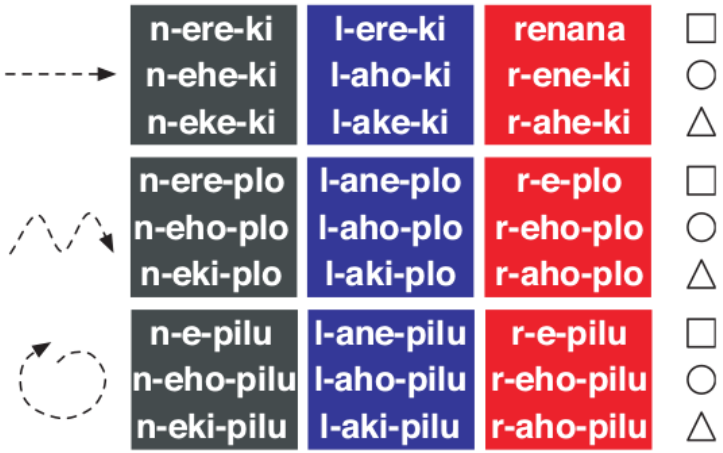
Kirby, Griffiths & Smith (2014). Iterated learning and the evolution of language.

Iterated Learning

"We show that languages transmitted culturally evolve in such a way as to maximize their own transmissibility: over time, the languages in our experiments become **easier to learn** and **increasingly structured**. Furthermore, this structure emerges purely as a consequence of the transmission of language over generations [...]"

Experiment 1 (Outcome)

Experiment 2 (Outcome)



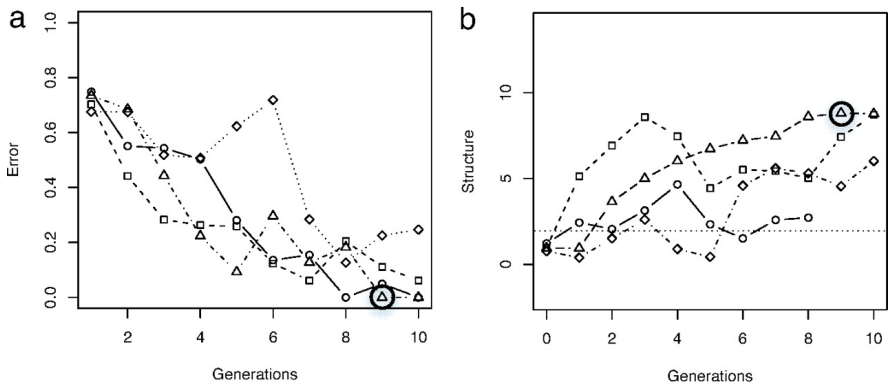
Kirby et al. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language.

Experiment 1: Learning Pressure

In Experiment 1, only **learning pressure** is given. That is, participants need to learn and remember string/meaning mappings. Their output on unseen mappings is handed down to the next generation. In this setup, the “languages” collaps into **underspecified** states.

Table 1. Number of distinct words by generation in the first experiment

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	17	9	6	5	4	4	2	2	2	2
□ Chain 2	27	17	15	8	7	6	6	6	5	5	4
△ Chain 3	27	24	8	6	6	5	6	5	5	5	5
◇ Chain 4	27	23	9	10	9	11	7	5	5	4	4



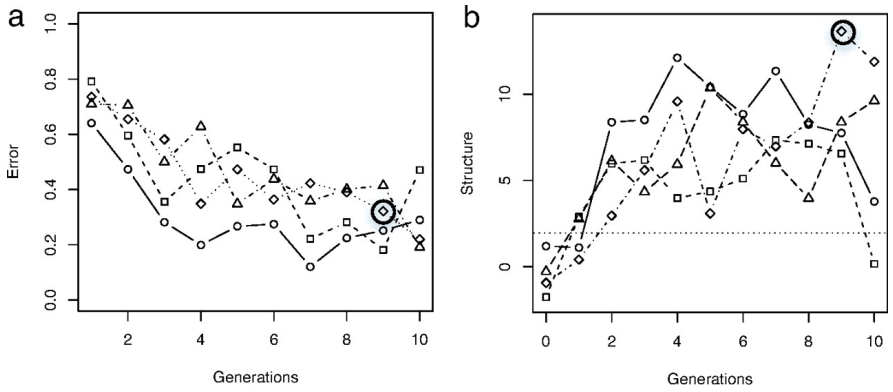
Note: The circled triangle marks the language which is depicted in the above slide on the left, i.e. Experiment 1 (Outcome).

Experiment 2: Learning + Communicative Pressure

In Experiment 2, “if any strings were assigned to more than 1 meaning, all but 1 of those meanings (chosen at random) was removed from the training data.” As a consequence, more **structured signals** emerge.

Table 2. Number of distinct words by generation in the second experiment

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	23	22	17	21	21	17	21	25	13	16
□ Chain 2	27	26	13	10	10	16	16	12	12	13	12
△ Chain 3	27	11	16	14	12	17	14	16	20	19	12
◇ Chain 4	27	19	19	17	19	17	22	23	21	27	23

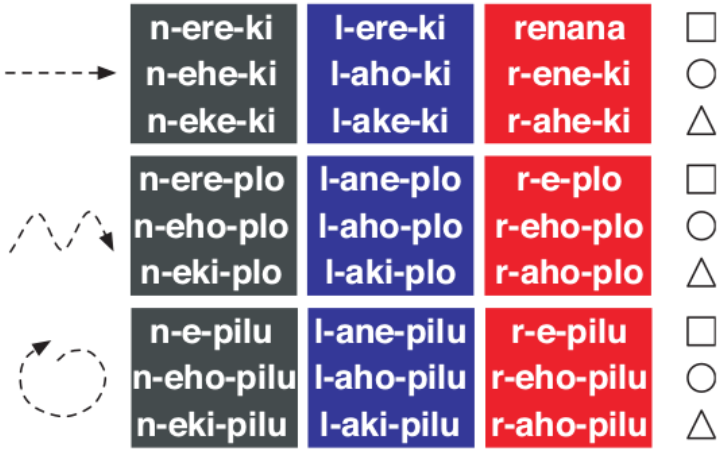


Exercise

Assume the outcome “languages” of Experiment 1 and 2 given below. Calculate the TTR for linguistic forms in each. What does the difference in TTR tell us about the relationship between *form* and *meaning* in these artificial languages? Could this insight be extended to “real world” languages?

Experiment 1 (Outcome)

Experiment 2 (Outcome)



Further Studies

Computational models

Kirby et al. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*.

Kirby and Hurford (2002). The emergence of linguistic structure: An overview of the iterated learning model. In: Cangelosi and Parisi (Eds.), *Simulating the evolution of language*.

Kirby (2001). Spontaneous evolution of linguistic structure – An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions of Evolutionary Computation*.

Graphical communication

Tamariz and Kirby (2015). Culture: copying, compression, and conventionality. *Cognitive Science*.

Sequence learning

Cornish et al. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PloS ONE*.

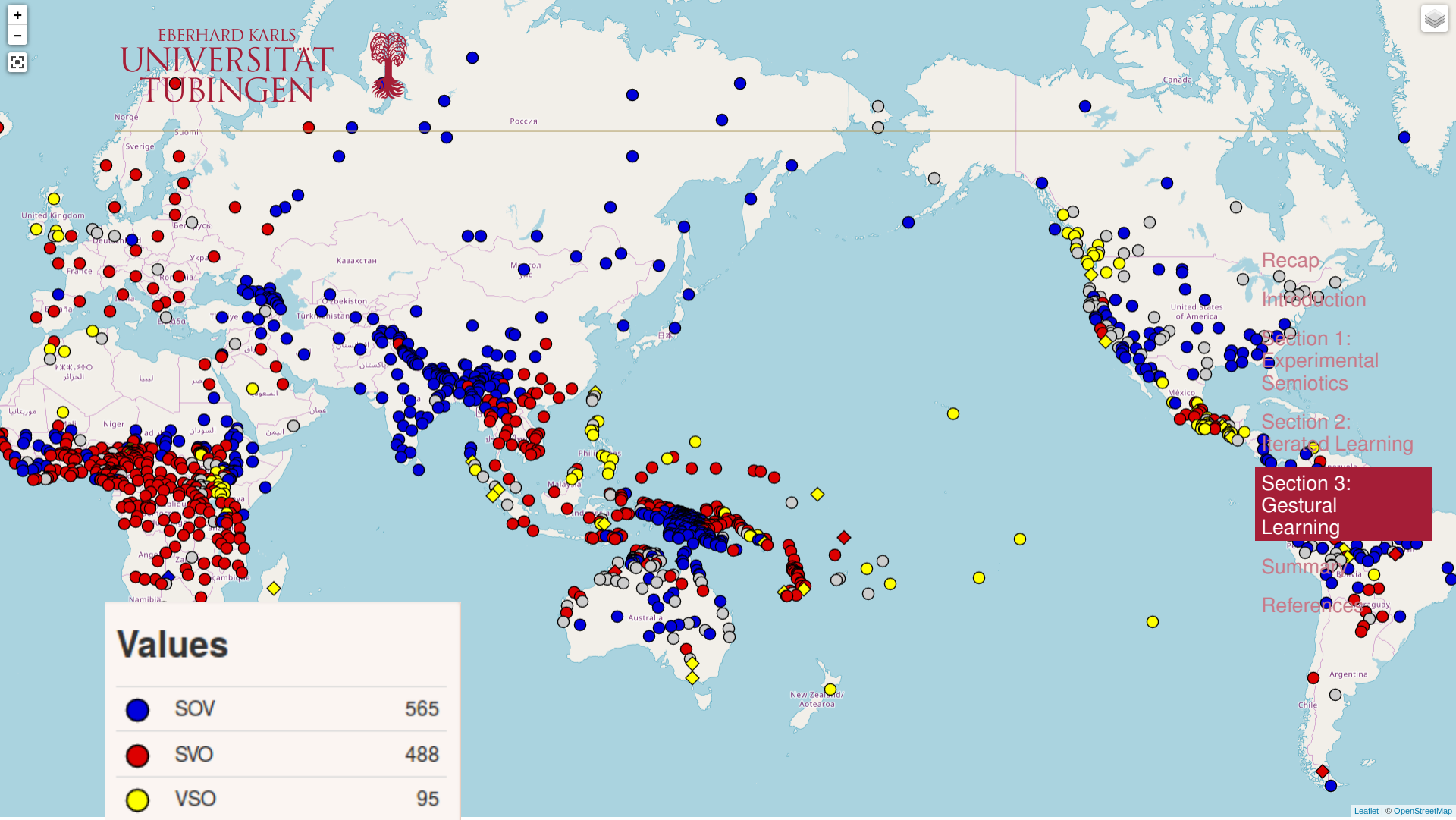
Artificial language learning

Smith et al. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transaction B*.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 3: Gestural Learning



Background: Word Order

Order of subject, object, and verb for 1377 languages in *World Atlas of Language Structures* (WALS).

Background: Word order in "Proto-World"?

PNAS

The origin and evolution of word order

Murray Gell-Mann[a,1] **and Merritt Ruhlen**[b,1]

[a]Santa Fe Institute, Santa Fe, NM 87501; and [b]Department of Anthropology, Stanford University, Stanford, CA 94305

Contributed by Murray Gell-Mann, August 26, 2011 (sent for review August 19, 2011)

Recent work in comparative linguistics suggests that all, or almost all, attested human languages may derive from a single earlier language. If that is so, then this language—like nearly all extant languages—most likely had a basic ordering of the subject (S), verb (V), and object (O) in a declarative sentence of the type "the man (S) killed (V) the bear (O)." When one compares the

man") and uses prepositions. (Nowadays, these correlations are described in terms of head-first and head-last constructions.) In light of such correlations it is often possible to discern relic traits, such as GN order in a language that has already changed its basic word order from SOV to SVO. Later work (7) has shown that diachronic pathways of grammaticalization often reveal relic

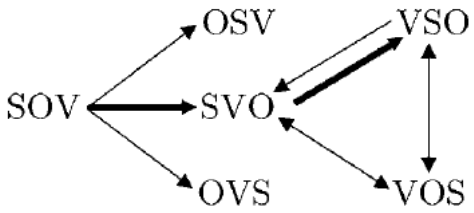


Fig. 1. Evolution of word order.

Gell-Mann, Murray & Merritt Ruhlen (2011). The origin and evolution of word order.

Basic Order of Events

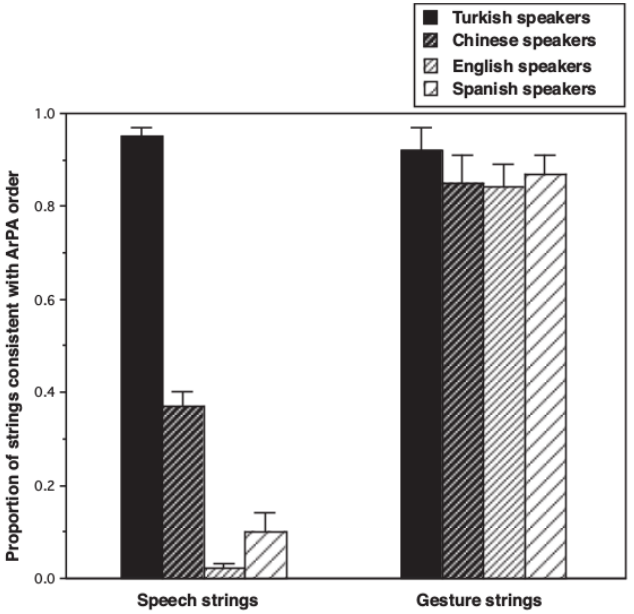


Fig. 2. Proportion of speech (*Left*) and gesture (*Right*) strings produced by speakers of Turkish, Chinese, English, and Spanish to describe transitive actions that were consistent with the ArPA order. Included are both in-place and crossing-space transitive actions.

Note: ArPA order means Actor (Ar)– Patient (P) – Action (A), i.e. SOV in language typology.

Goldin-Meadow et al. (2008). The natural order of events: How speakers of different languages represent events nonverbally.

Semantics and Word Order: Extensional and Intensional Events

extensional (above), intensional (below)

Fig. 1. Example item: intensional event. 'Pirate throws guitar'.

Fig. 2. Example item: extensional event. 'Cook thinks of sock'.

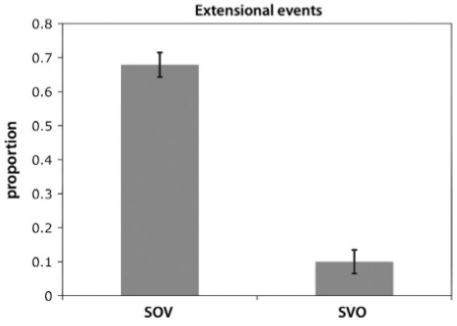


Fig. 3. Results: extensional events. Mean proportions of SOV and SVO gesturing orders for extensional events. Error bars indicate standard error of the mean.

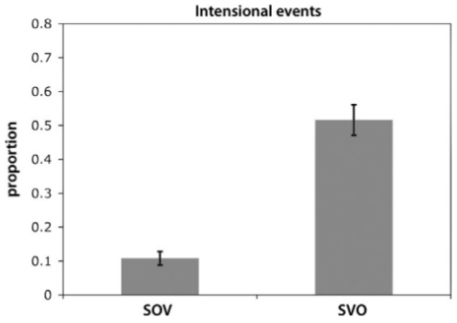
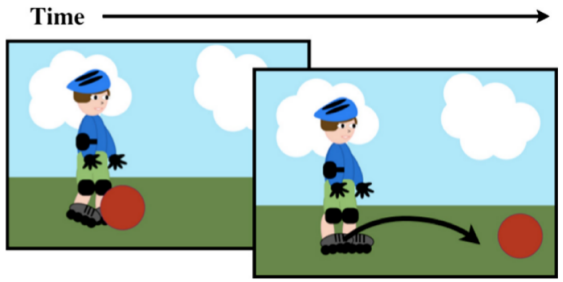


Fig. 4. Results: intensional events. Mean proportions of SOV and SVO gesturing orders for intensional events. Error bars indicate standard error of the mean.

Schouwstra, Marieke, and De Swart, Henriette (2014) The semantic origins of word order.

Semantics and Word Order: Reversible and Nonreversible Events

Nonreversible event



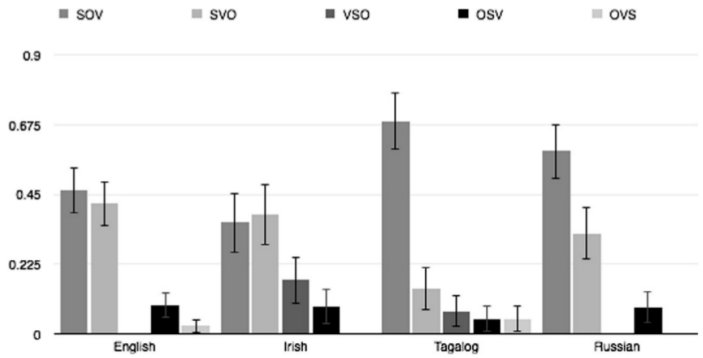


Fig. 2. Proportions of responses in basic three-word orders for nonreversible events (inanimate objects).

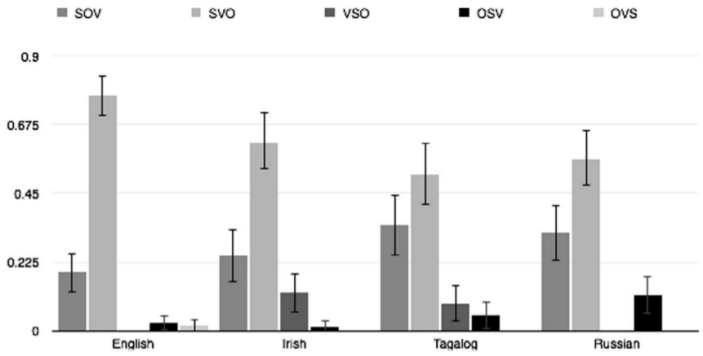


Fig. 3. Proportions of responses in basic three-word orders for reversible events (human objects).

Futrell et al. (2015) Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages.


photo by Abdullah Mohiuddin
CC BY-NC-SA Hajime Yamauchi
GGCCTGACACCTTCGCACC

Summary

Summary

- **Semiotics** deals with **signs** and their meanings.
- Signs are often further subdivided into **icons**, **indeces**, and **symbols**.
- **Experimental semiotics** is a framework to model language/symbolic evolution in the lab.
- Some of the main types of experiments include: **artificial language learning**, **iterated learning**, **dyadic interaction**.
- One crucial observation is that symbols evolve from icons/indeces via **repetition**, **interaction**, and hence **conventionalization**.


photo by Abdullah Mohiuddin
CC BY-NC-SA Hajime Yamauchi
GGCCTGACACCTTCGCACC

References

References

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., and Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, 136, 215-221.

Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, 31(6), 961-987.

Gell-Mann, M., and Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, 108(42), 17290-17295.

Goldin-Meadow, S., So, W. C., Özyürek, A., and Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, 105(27), 9163-9168.

Hockett, Charles F. (1960). The origin of speech. *Scientific American*, Vol. 203, No. 3, pp. 88-97.

Kirby, S., Griffiths, T., and Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108-114.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681-10686.

Kroeger, Paul (2019). *Analyzing meaning: An introduction to semantics and pragmatics. Second corrected and slightly revised edition.* (Textbooks in Language Sciences 5). Berlin: Language Science Press.

Little et al. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*.

Nölle, J., and Galantucci, B. (2022). Experimental semiotics: past, present, and future. In: The Routledge Handbook of Semiosis and the Brain.

Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning*. Orlando, Florida: Harcourt Brace Jovanovich.

Peirce, C. S. (2016). *Prolegomena to a science of reasoning. Phaneroscopy, semeiotic, logic*. Peter Lang Verlag.

Schouwstra, M., and de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3), 431-436.

Thank You.

Contact:

Faculty of Philosophy
General Linguistics
Dr. Christian Bentz
SFS Keplerstraße 2, Room 168
chris@christianbentz.de
Office hours:
During term: Wednesdays 10-11am
Out of term: arrange via e-mail