word co-activation and competition and the perception of morphological structure

VITO PIRRELLI - Comphys Lab
INSTITUTE FOR COMPUTATIONAL LINGUISTICS, PISA CNR ITALY

ComPhys
physiology of communication

the lexicon as a dynamic system

the linguistic viewpoint (Blevins 2006)

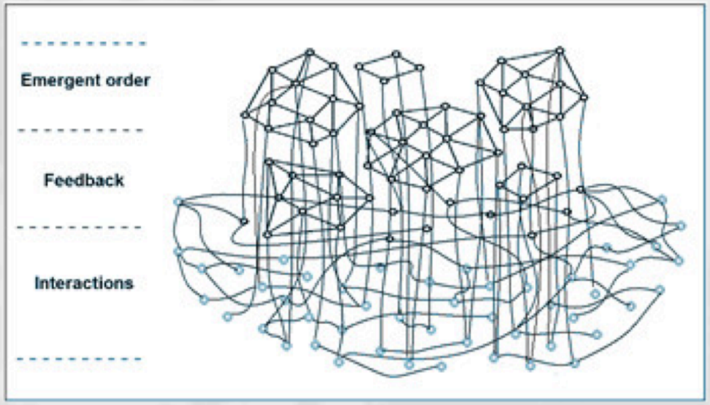
- the lexicon as a dynamic system imposes an abstractive view of morphological competence, whereby sublexical units are the by-product of a process of **self-organisation** of full forms
- such units are **not epiphenomenal**, but they affect the way speakers access, activate and recall lexical items in the LM memory, shaping up their **expectations**
- this architecture reduces the computational complexity of the problem and change its nature considerably
  - there are no distinct input and output representation levels
  - there is no way to dichotomise representations from operations defined over representations
  - basic processing mechanisms required: **co-activation, competition, selection**

BLEVINS, JAMES. 2006. Word-based morphology. Journal of Linguistics 42. 531-573.

compounding

rationale: investigate dynamics of word production

- "idea" ➔ lemma retrieval ➔ implementation of the **motor plan**
- is production **sequential** or **cascaded**?
  - cognitive processes do not influence motor execution (e.g., Damien 2003)
  - cognitive processes might influence motor execution (e.g., Delattre, Bonin & Barry 2006)



- what is the input to the motor system? words or smaller units such as morphemes?

Gagné, C.L., & Spalding, T.L. (in press). Effects of morphology and semantic transparency on typing latencies in English compound and pseudo-compound words. Journal of Experimental Psychology: Learning, Memory, and Cognition.

design

COMPOUND		N
TT	*blueberry*	*59*
TO	*honeycomb*	*53*
OT	*chopstick*	*46*
OO	*hogwash*	*42*
		200
MONOMORPHEMIC		
	gradient	*200*

 # letters: Mean = 8.6 range 6-11

PSEUDO COMPOUND	N
carpet	*50*
MONOMORPHEMIC	
gradient	*50*

 # letters: Mean = 6.7 range 6-9

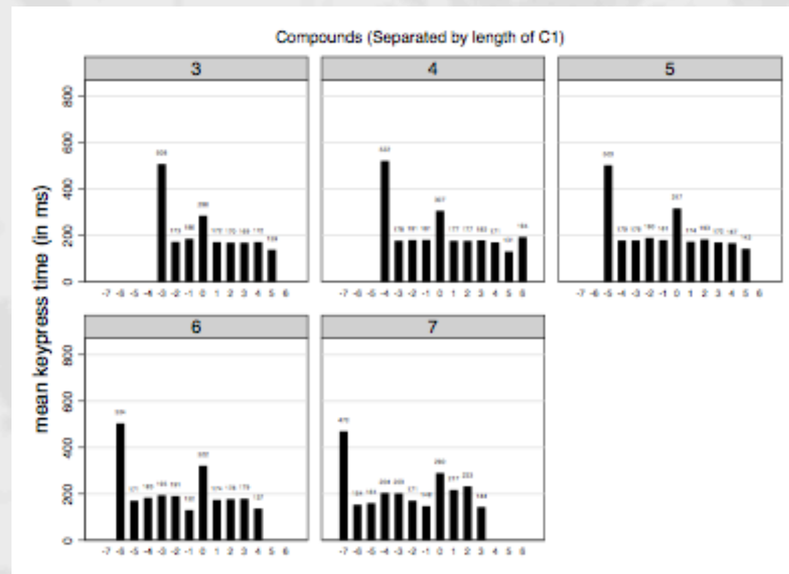
divided the set of 500 words into two lists.
Each participant saw one list.

recording typing time

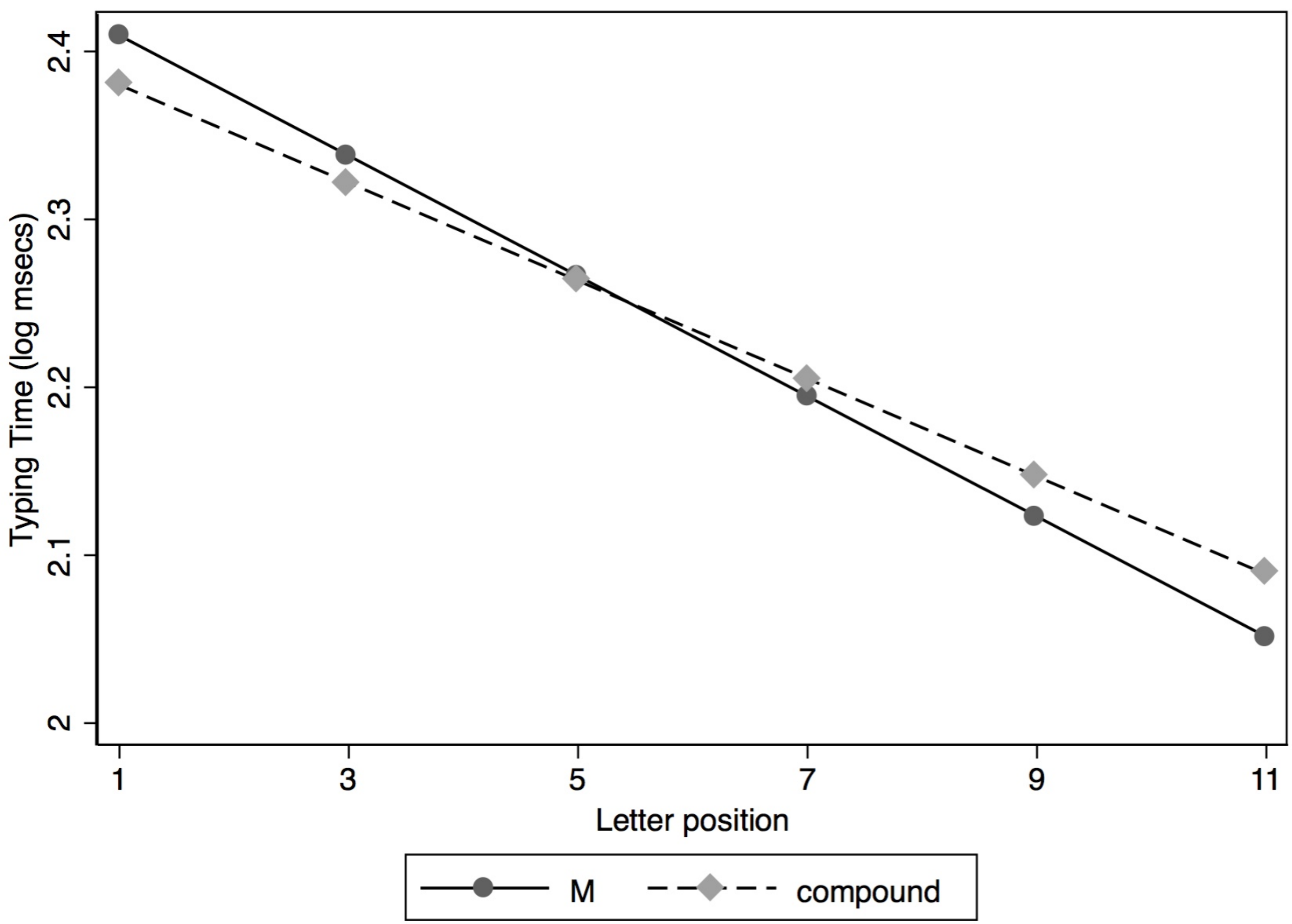
- progressive demasking, then typing
- record the typing time for each letter
- **elevation** in typing latency for the initial letter of the 2nd constituent; morphemic representations function as planning units
- data analyses with linear mixed effects models, with subjects and items as random effects

s n o w b a l l

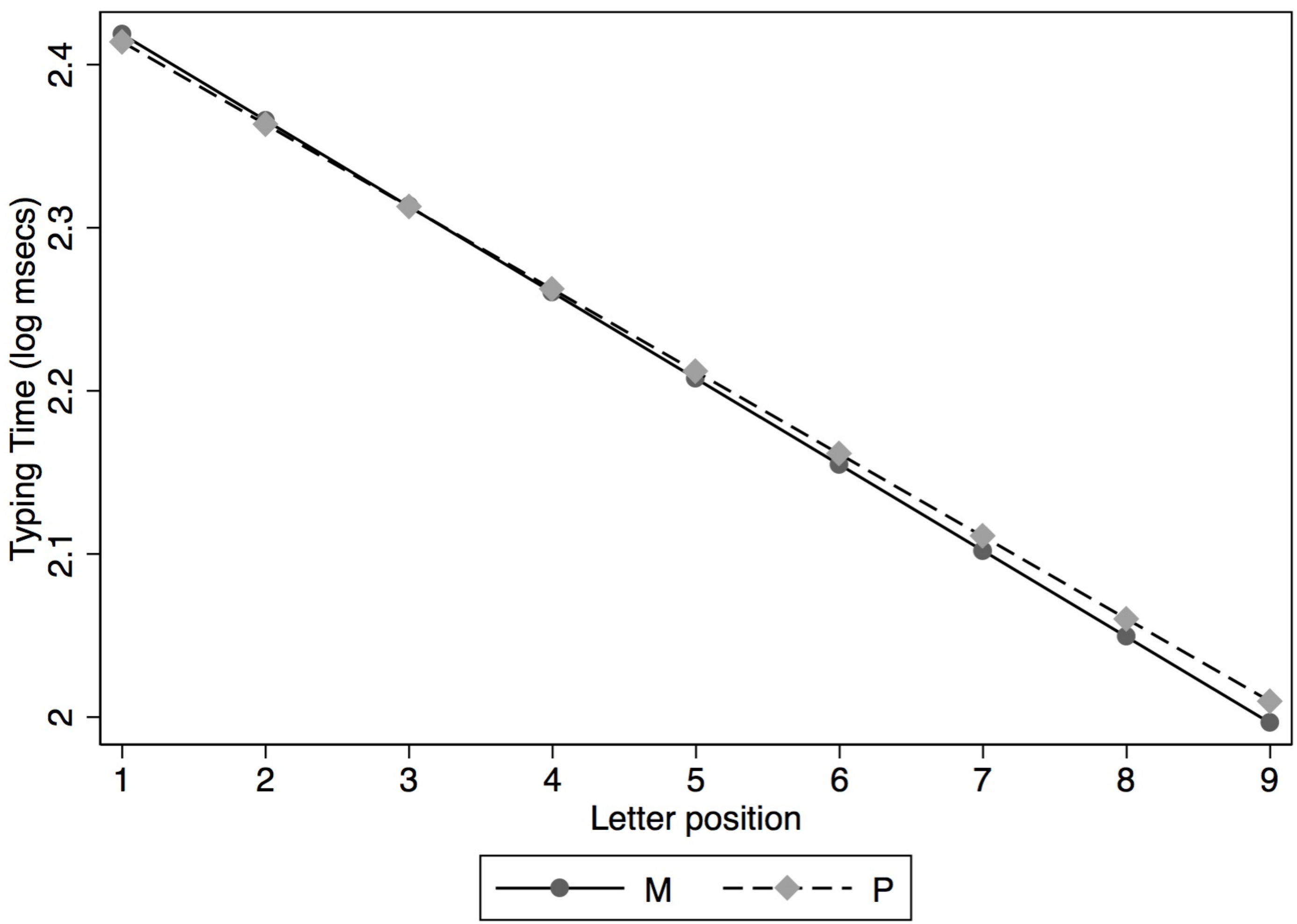
s n o w b a l l



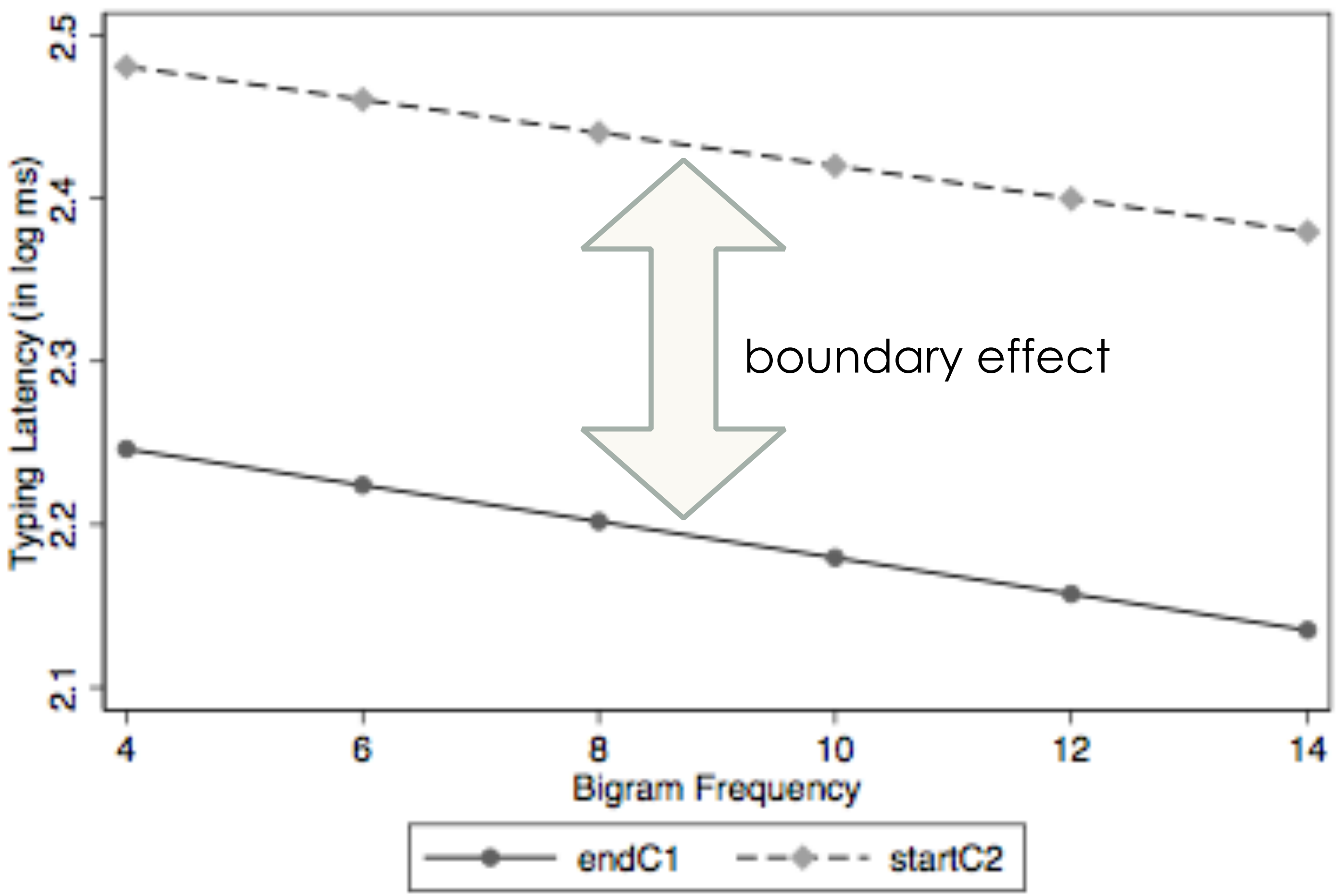
speedup is faster for monomorphemic words than for compounds



speedup is **not** faster for monomorphemic words than for pseudos



influence of bigram frequency

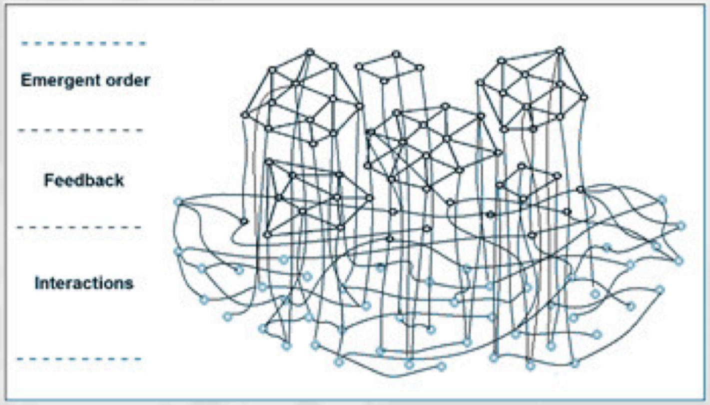


human evidence

- **morphemes act as planning units**, even for opaque compounds
- **semantic transparency of C1 and of C2 act differently**
  - influence of the 1st constituent's transparency but not of the 2nd constituent's
  - increased transparency of C1
    - slows production across whole word (due to interference e.g., *snow* and *snowball* both active)
- pseudocompounds are not fully equivalent to monomorphemic words: **embedded pseudo-morphemes** influence production

computer simulation: temporal memories for time-series

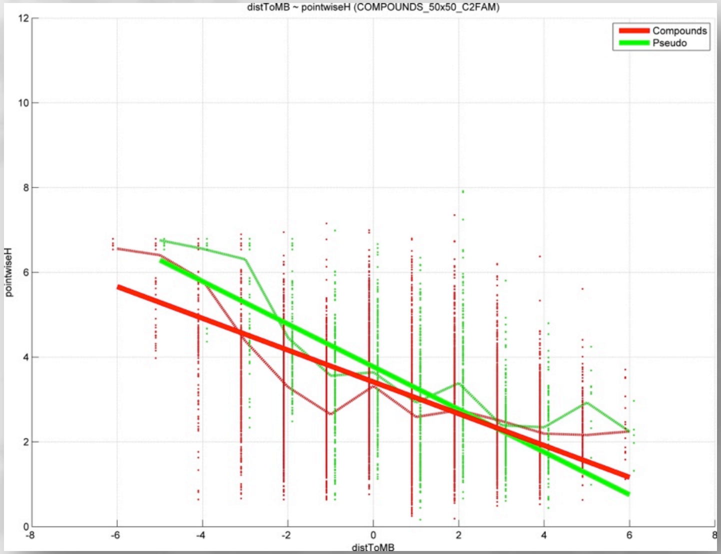
- use basic principles of **correlative learning**, accounting for the behaviour of **temporal maps** (TSOMs), to simulate the **emergence** of structure sensitivity in compound processing
- do these principles capture relevant distributional effects in input data?
- at what representation levels (letter, morpheme, word) are these effects observed?
- do they correlate with human evidence?



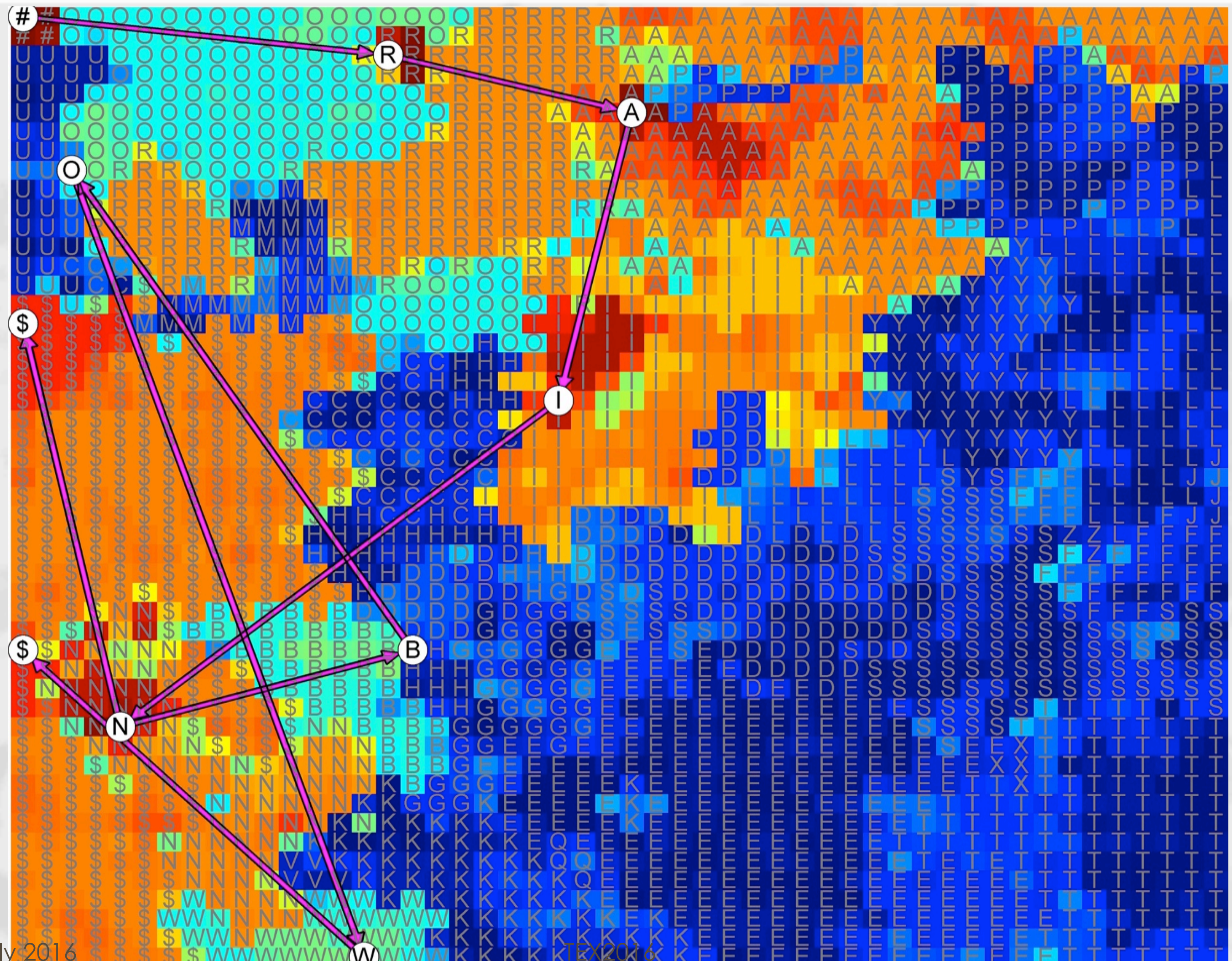
M. Ferro, C. Marzi, V. Pirrelli , C. Gagné, T. Spalding 2016. One word or two? Discriminative effects of word entrenchment and competition on processing compounds and pseudo-compounds. IMM 17. Vienna.

design

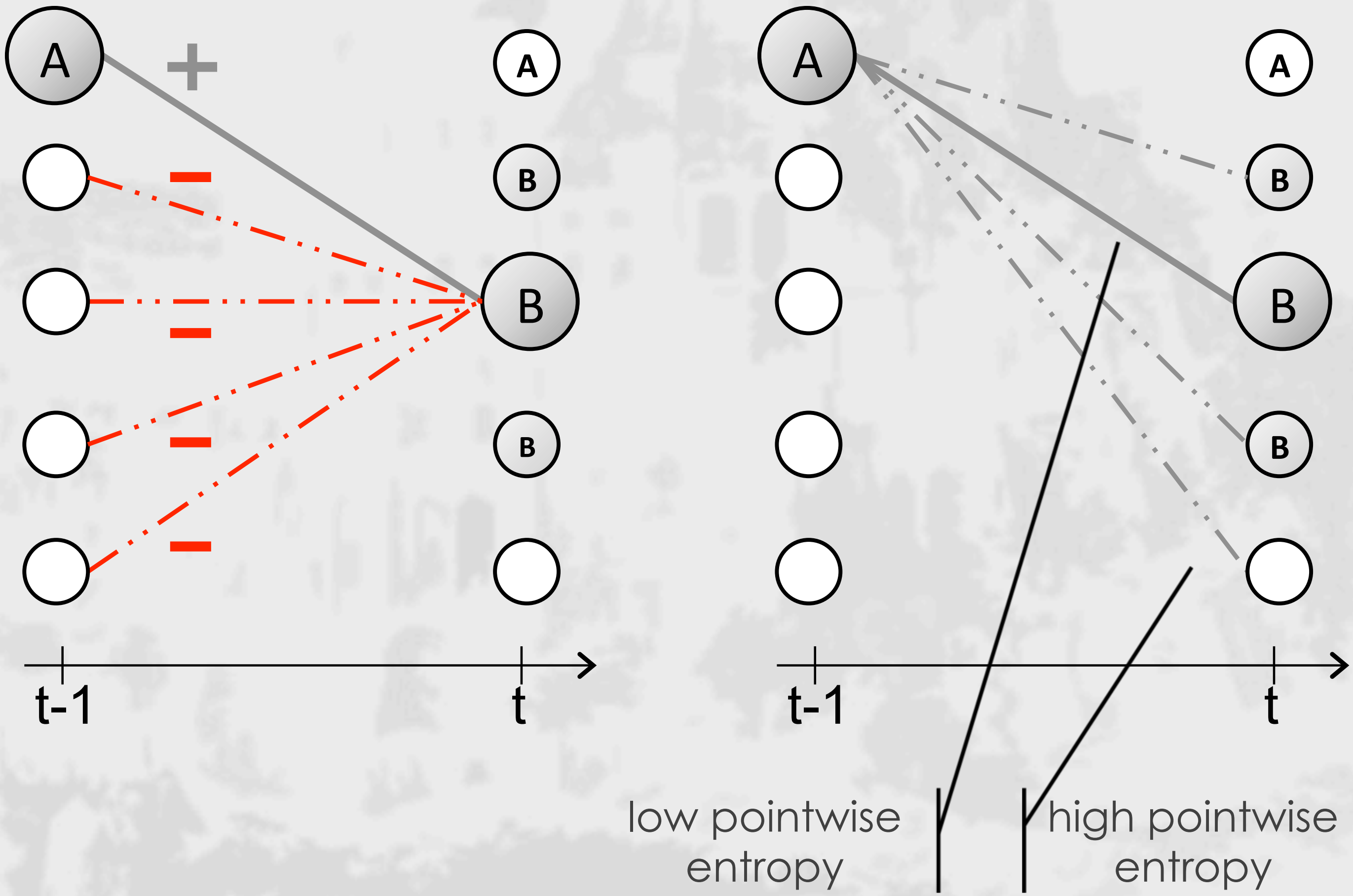
- basic training set
  - 200 compounds
  - 50 pseudo-compounds
- training protocol
  - vary input frequency of constituents and pseudocompounds, while keeping compound frequency fixed
  - for each regime, repeat experiment 5 times
  - monitor the influence of different frequency distributions on perception of structure
- augmented training set
  - add to basic set compound families (on both C1 and C2)
  - obaserve the influence of compound families on perception of structure
- data analysis with linear mixed effects models, with repetitions and items as random effects



computer simulation: temporal memories for time-series



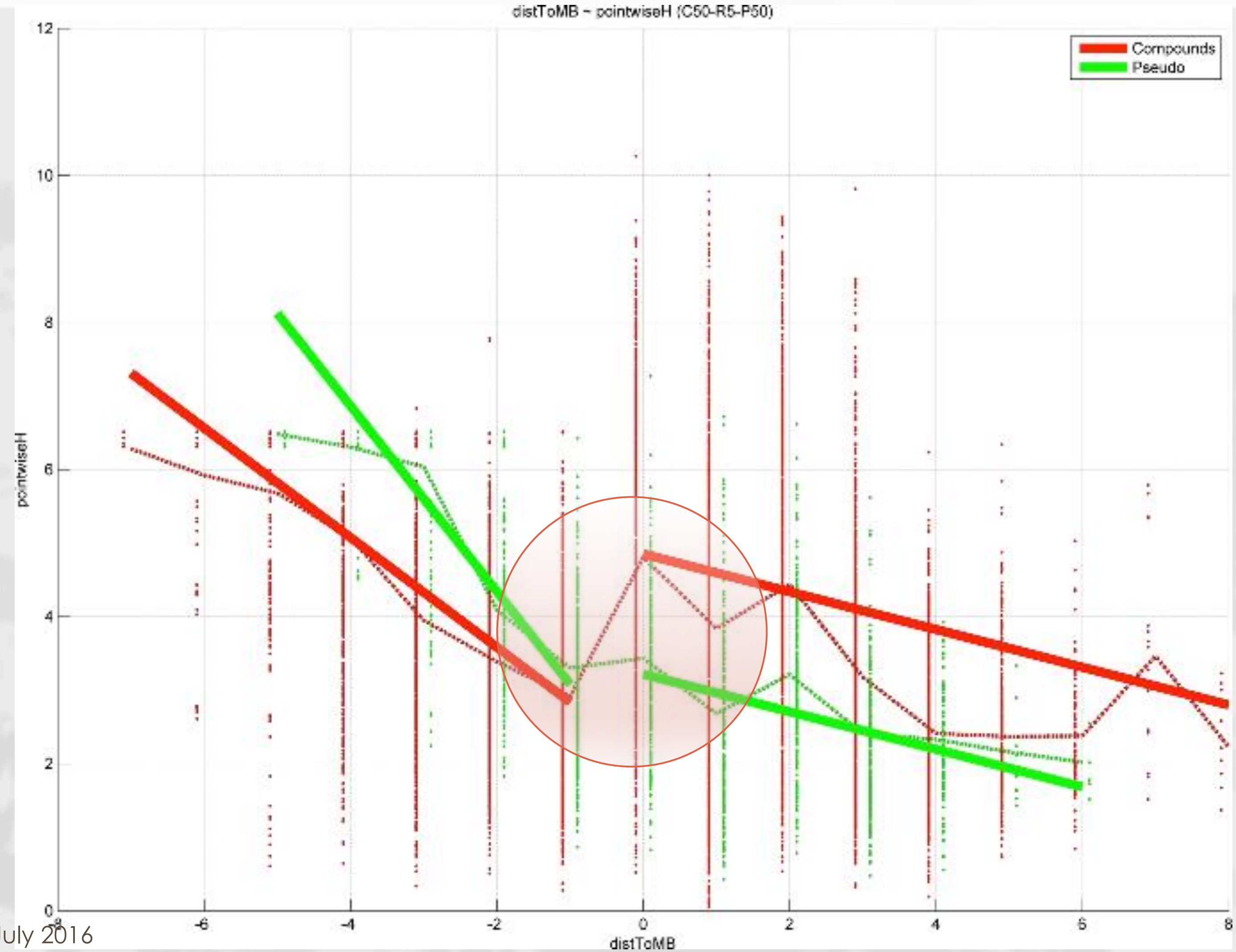
Hebbian learning



low pointwise entropy

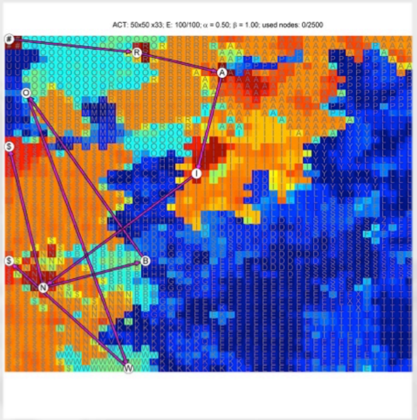
high pointwise entropy

pointwise entropy at morpheme boundary



simulation results

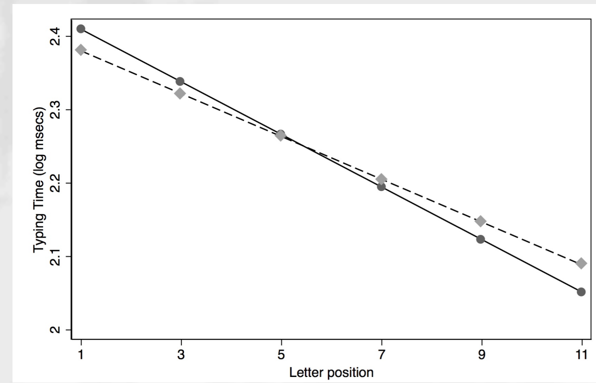
- distinct processing effects in pseudos and compounds
- rate of speedup faster in pseudos than in compounds (similar to monorphemic words in humans)
- position-independent bigram frequency **does not** significantly influence surprise at morpheme boundary
- but fine-grained context-sensitive probalistic letter distributions **significantly** affect expectations and surprise for upcoming symbols in the map
- at the word level, **compound families** make structural discontinuinities at morpheme boundary more salient



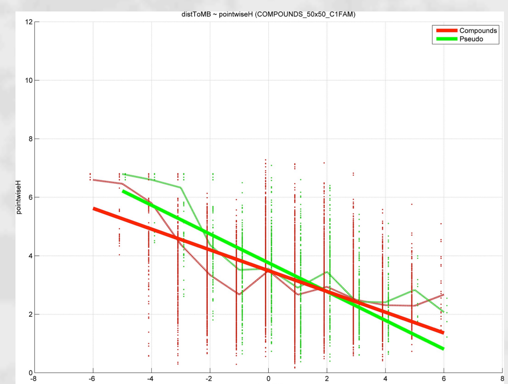
one experiment or two?

the two experiments address different **processing mechanisms**

- **human evidence**
  - cascaded models of word production
  - influence of semantic transparency on written production of compounds
  - influence of embedded pseudo-morphemes



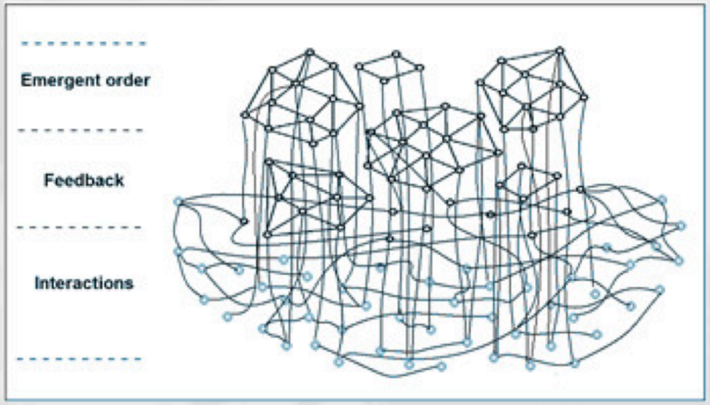
- **TSOM-based simulations**
  - peripheral levels of routinised recognition/production patterns
  - sensitivity to patterns of similarity and (dis)continuity in orthotactic/phonotactic time series
  - expectations & surprise over upcoming symbols reflect probability distributions in input data
  - purely morphological structures emerging from the probabilistic support provided by recurrent patterns in the input, at both letter level and constituent level



paradigm acquisition

rationale: investigate dynamics of paradigm acquisition

- "idea" ➔ words acquired in "families" ➔ family structure influences **pace of acquisition**



- does **frequency** help?
  - frequency by regularity interaction

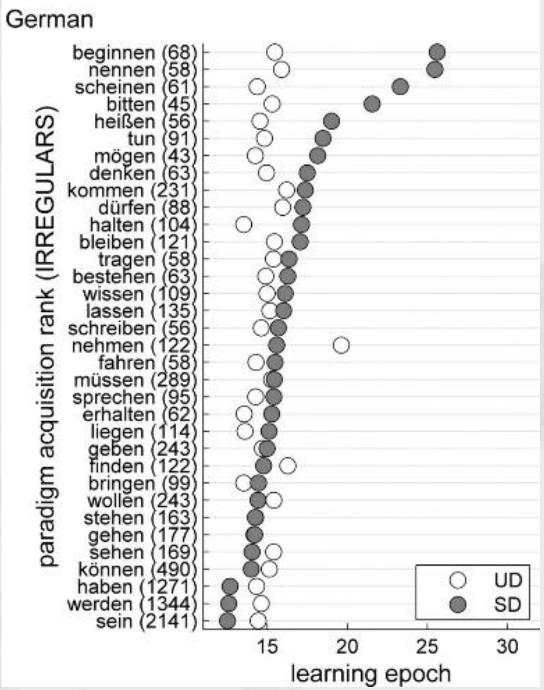
- what is the input to acquisition? paradigms, words or smaller units such as morphemes?

C. Marzi, M. Ferro, V. Pirrelli. 2014. Morphological structure through lexical parsability. LINGUE E LINGUAGGIO, XIII, vol. 2, p. 263-290, Bologna: il Mulino, ISSN: 1720-933.

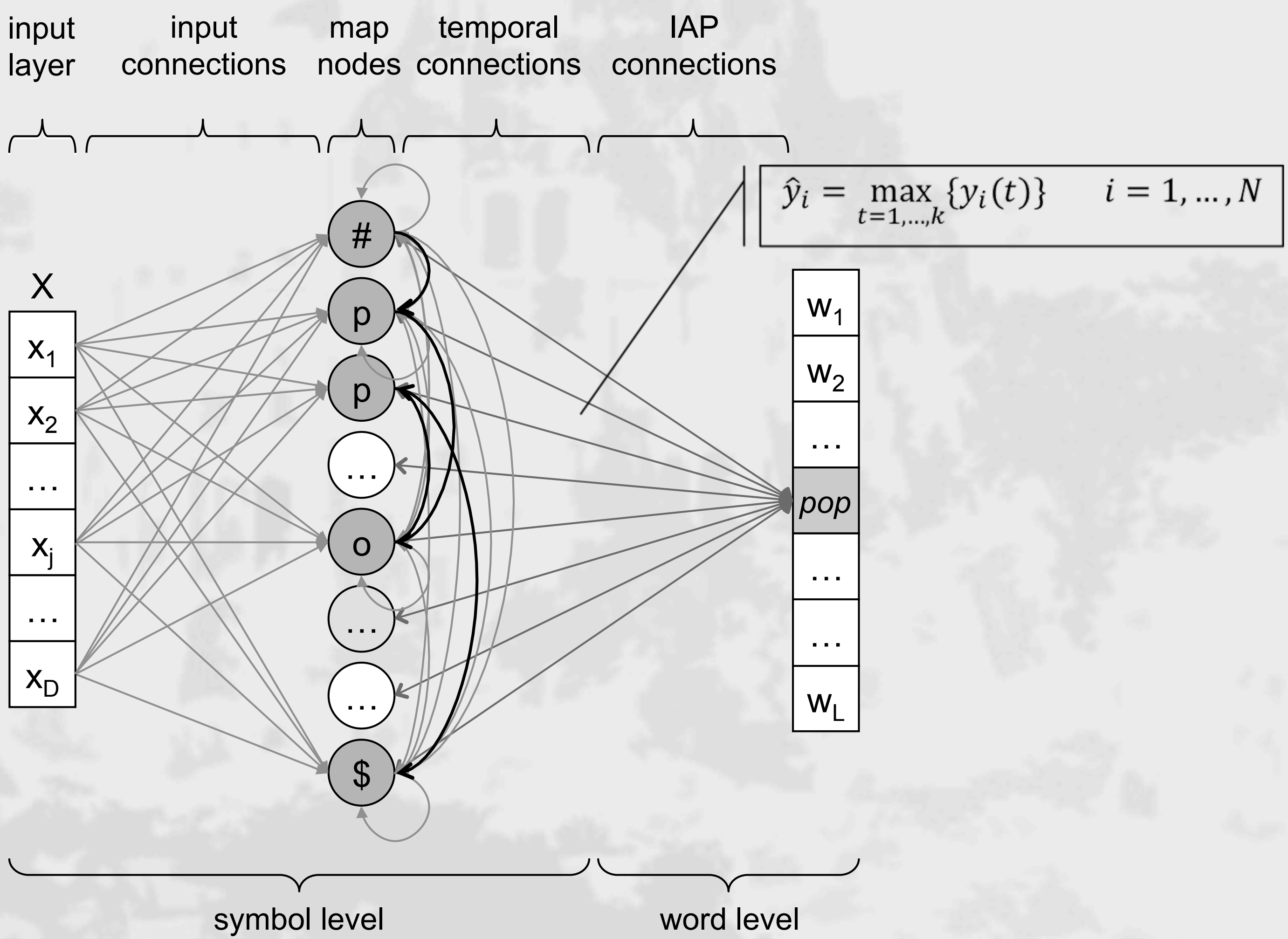
C. Marzi, F.A. Cardillo, M. Ferro, V. Pirrelli. 2016. Effects of frequency and regularity in an integrative model of word storage and processing. ITALIAN JOURNAL OF LINGUISTICS, 28, vol. 1.

design

- two languages
  - German
  - Italian
- basic training set
  - 50 top-frequency verb paradigms (both regular and irregular ones, selected by cumulative token freq)
  - 15 forms for each paradigm (from a fixed set of cells)
- two training regimes
  - uniform distributions: each form presented 5 times
  - realistic distributions (sampled from Celex and Paisà Italian Corpus)
- for each regime, repeat experiment 5 times
- test trained maps on recall accuracy
  - recoding accuracy
  - recall accuracy
- plot epoch of acquisition
  - by word form
  - by paradigm

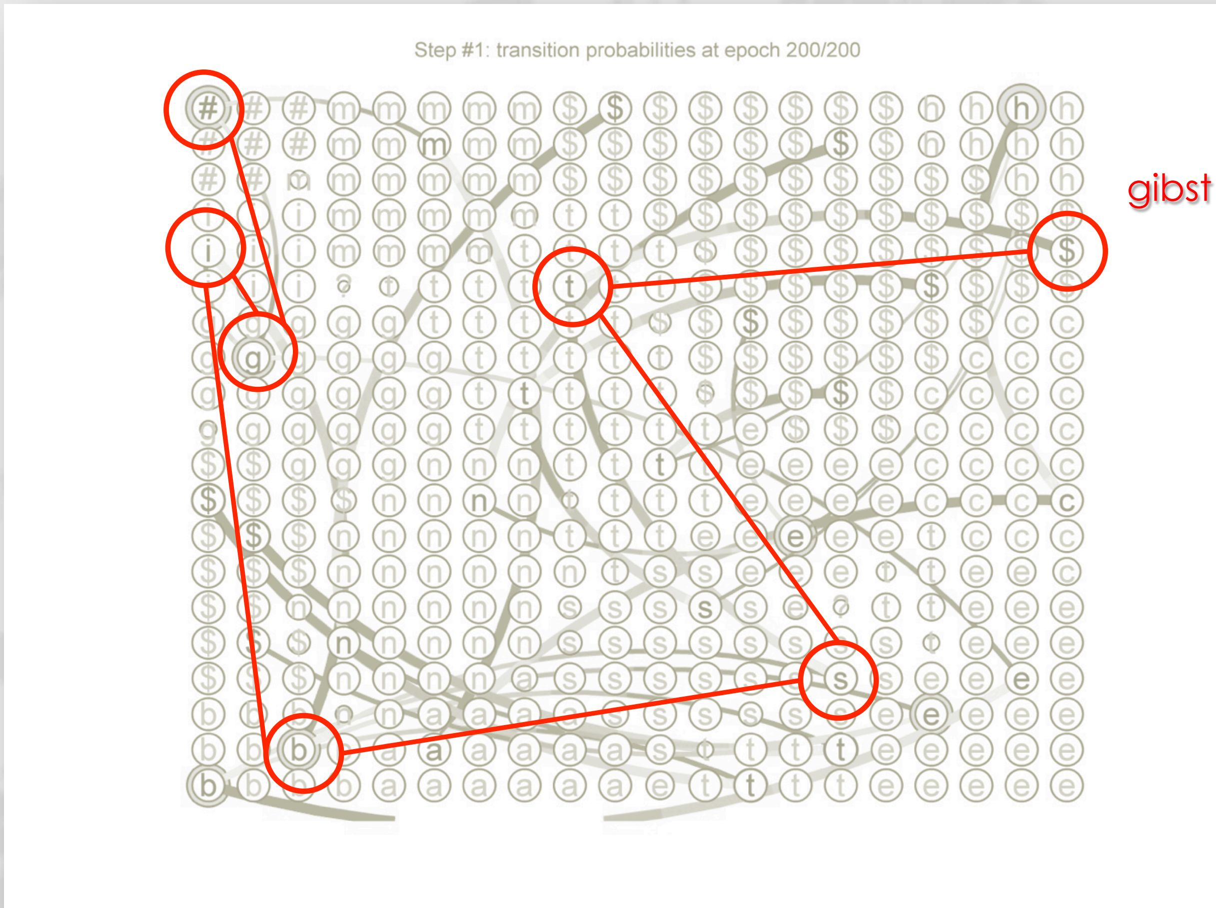


Integrated activation pattern (IAP)



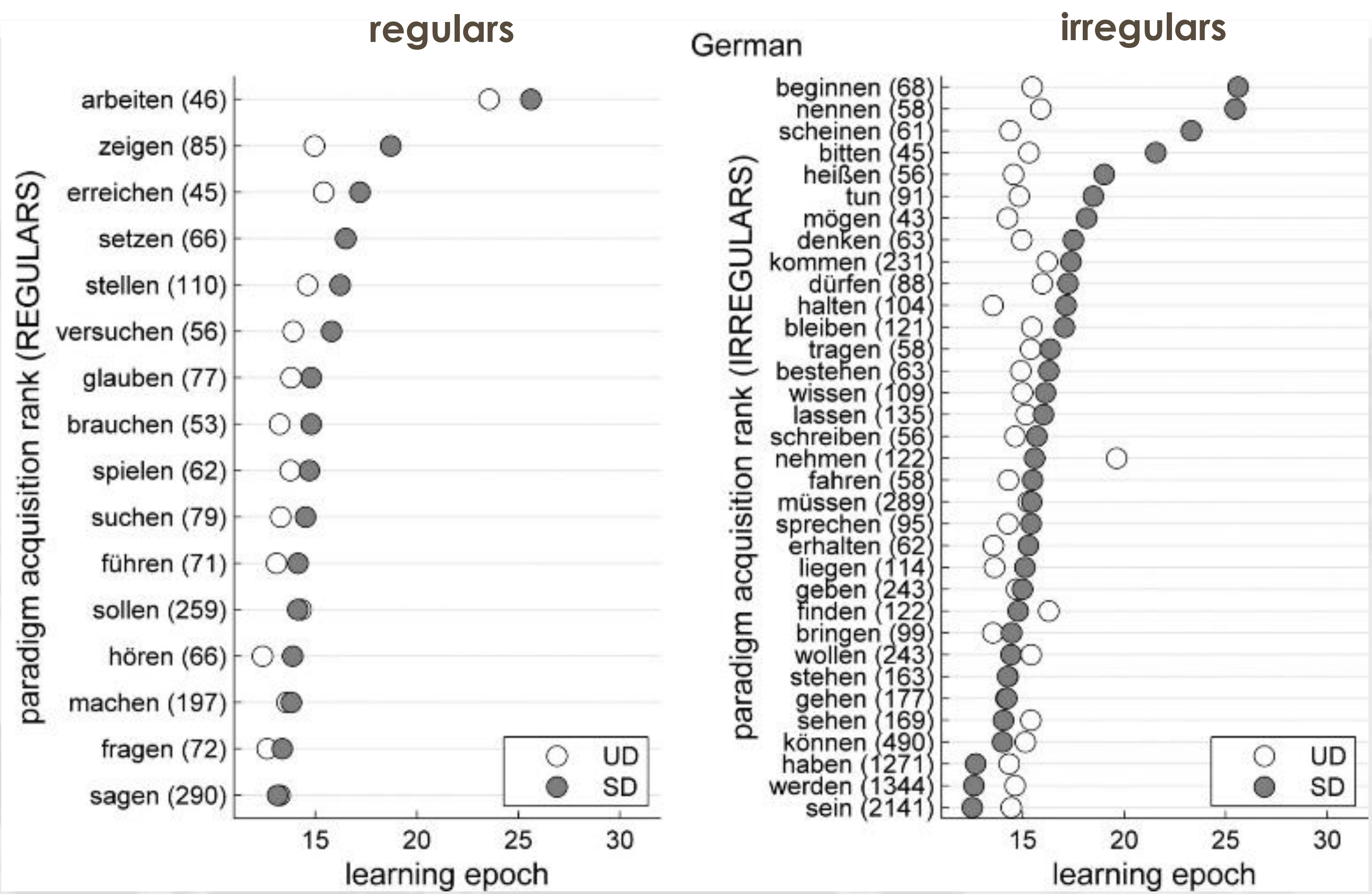
word recall from IAP

#

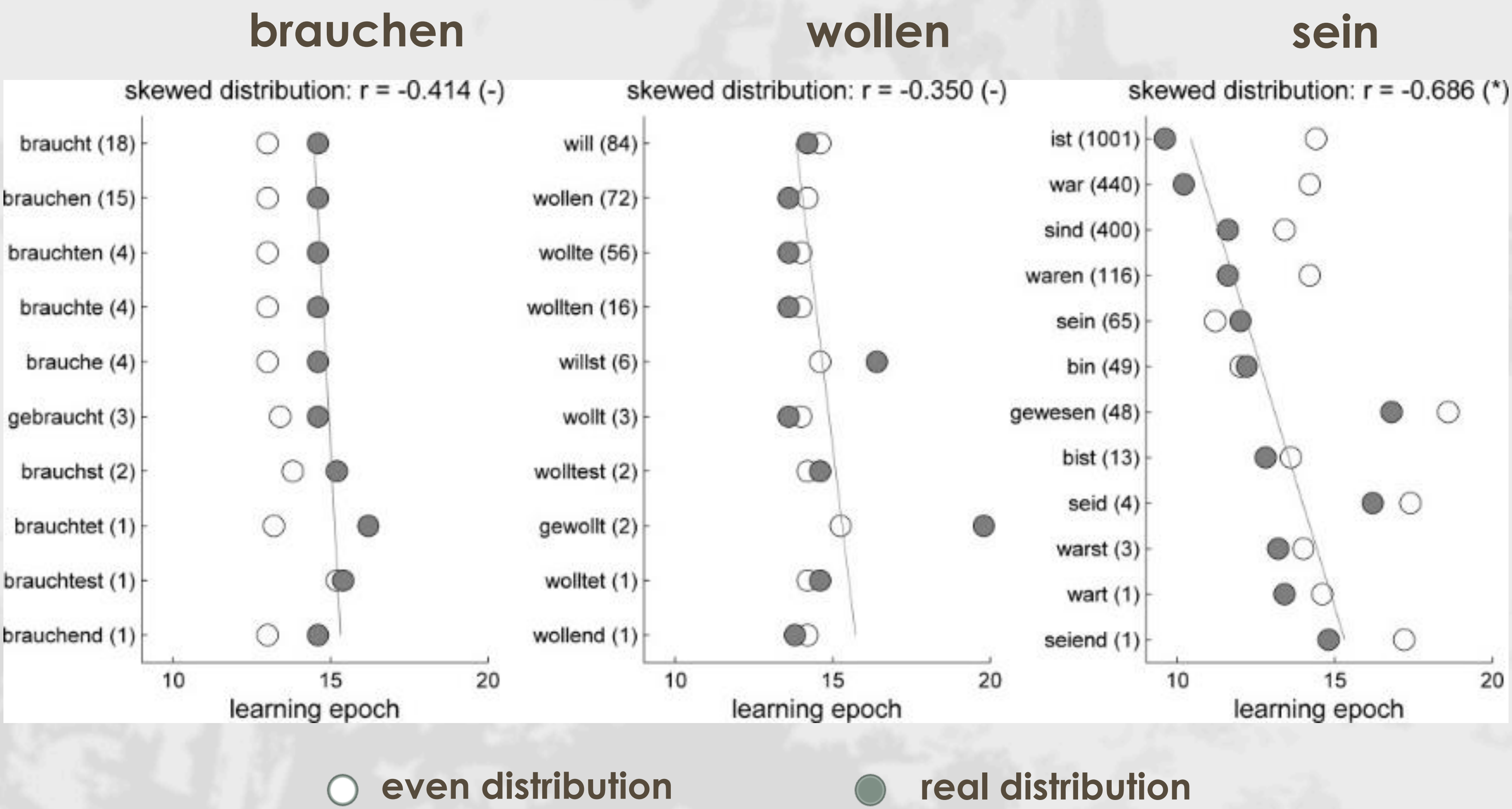


gibst

general picture

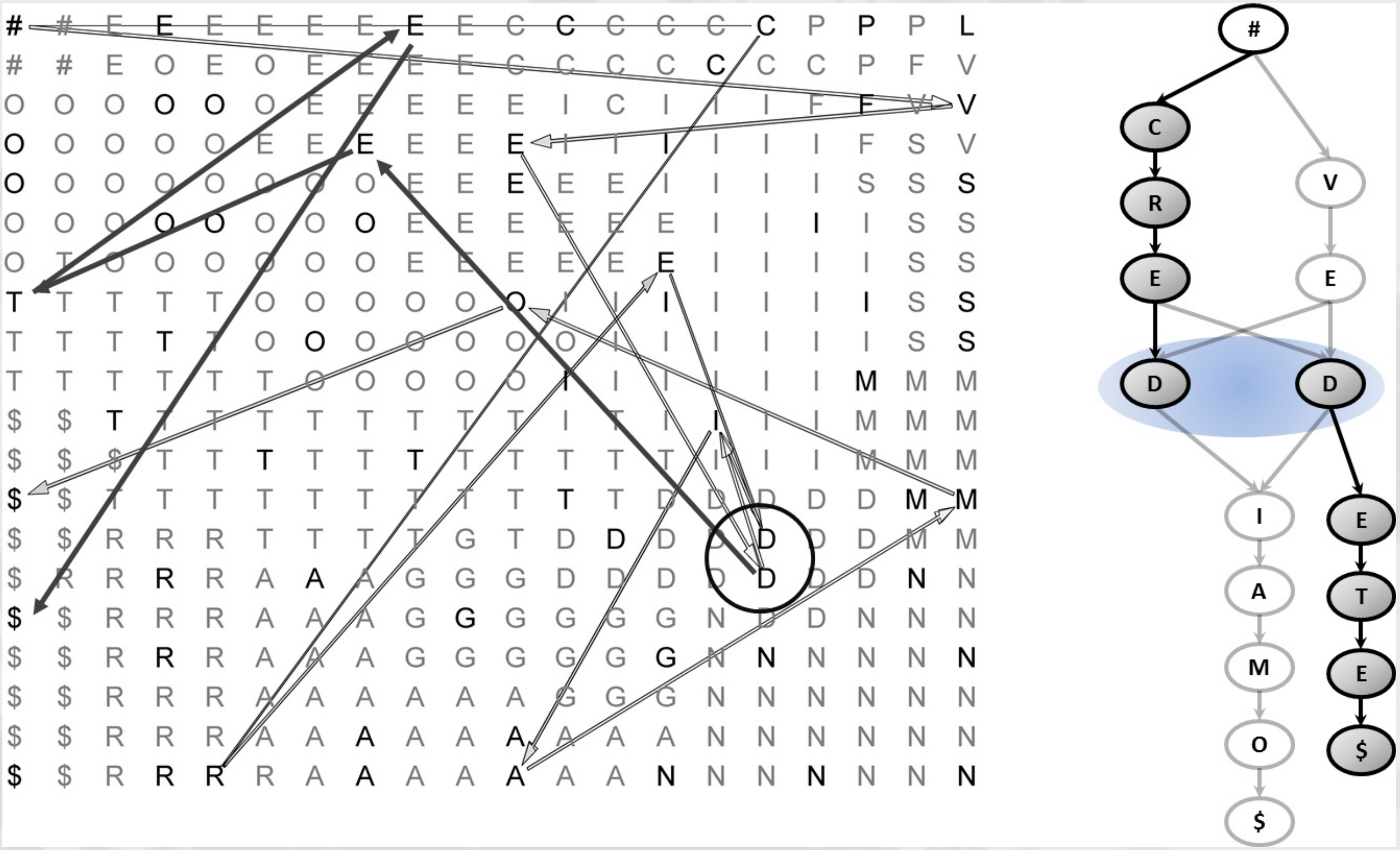


3 paradigms

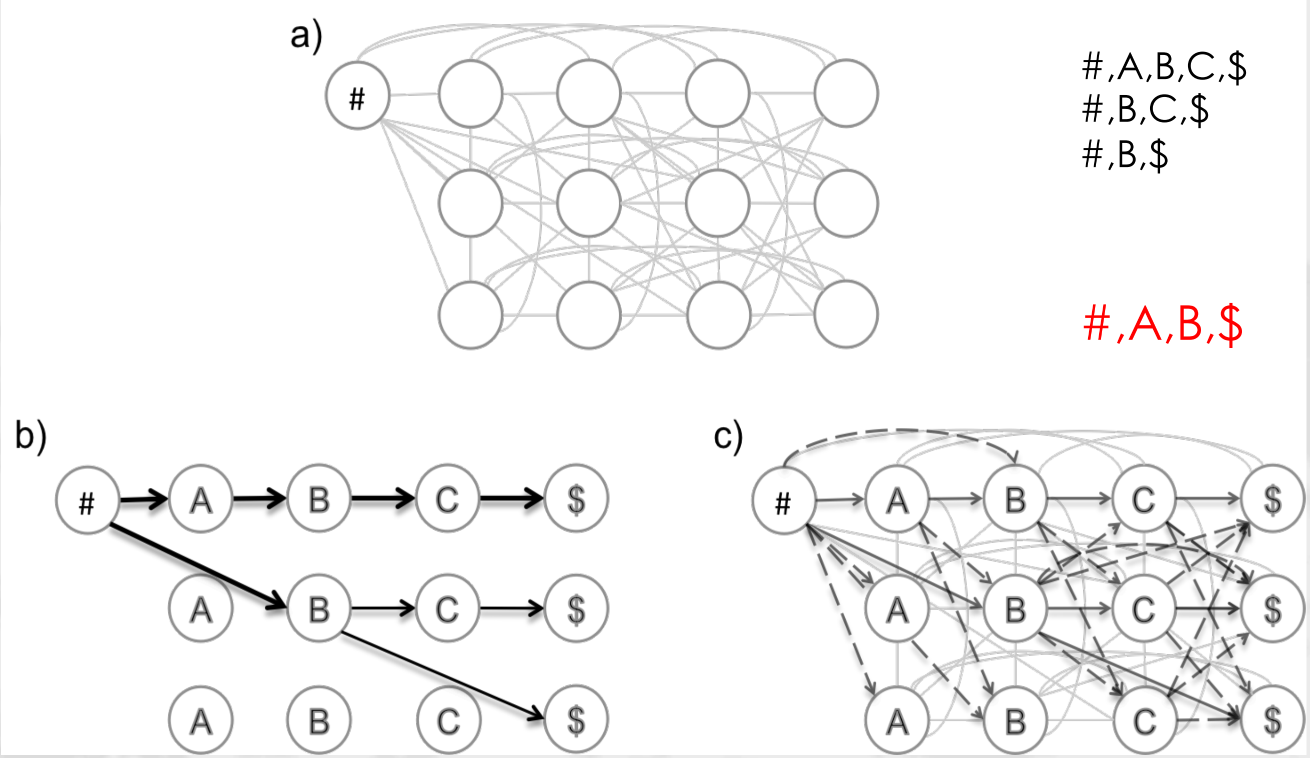


brauchen
wollen
sein
skewed distribution: r = -0.414 (-)
braucht (18)
brauchen (15)
brauchten (4)
brauchte (4)
brauche (4)
gebraucht (3)
brauchst (2)
brauchtet (1)
brauchtest (1)
brauchend (1)
10
15
20
learning epoch
skewed distribution: r = -0.350 (-)
will (84)
wollen (72)
wollte (56)
wollten (16)
willst (6)
wollt (3)
wolltest (2)
gewollt (2)
wolltet (1)
wollend (1)
10
15
20
learning epoch
skewed distribution: r = -0.686 (*)
ist (1001)
war (440)
sind (400)
waren (116)
sein (65)
bin (49)
gewesen (48)
bist (13)
seid (4)
warst (3)
wart (1)
seiend (1)
10
15
20
learning epoch
even distribution
real distribution

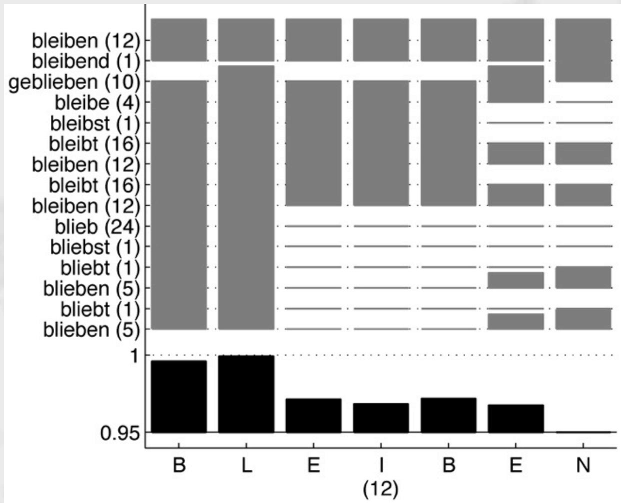
structure sharing

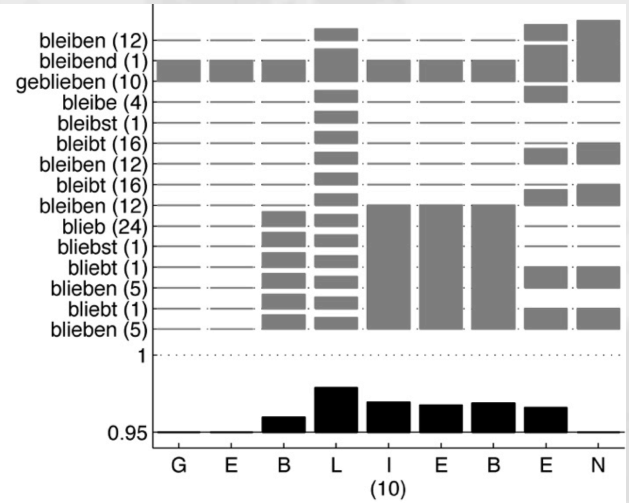


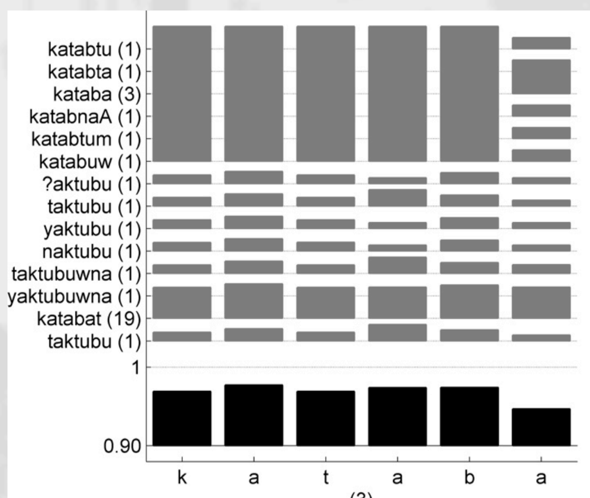
blending



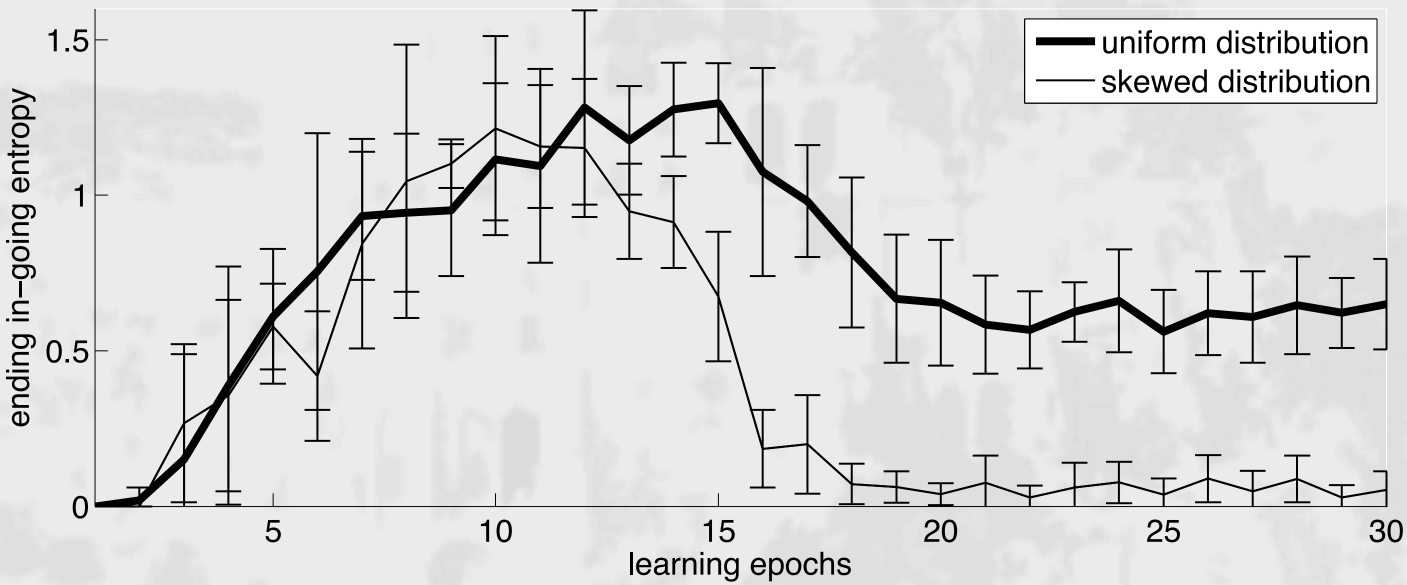
more on blending





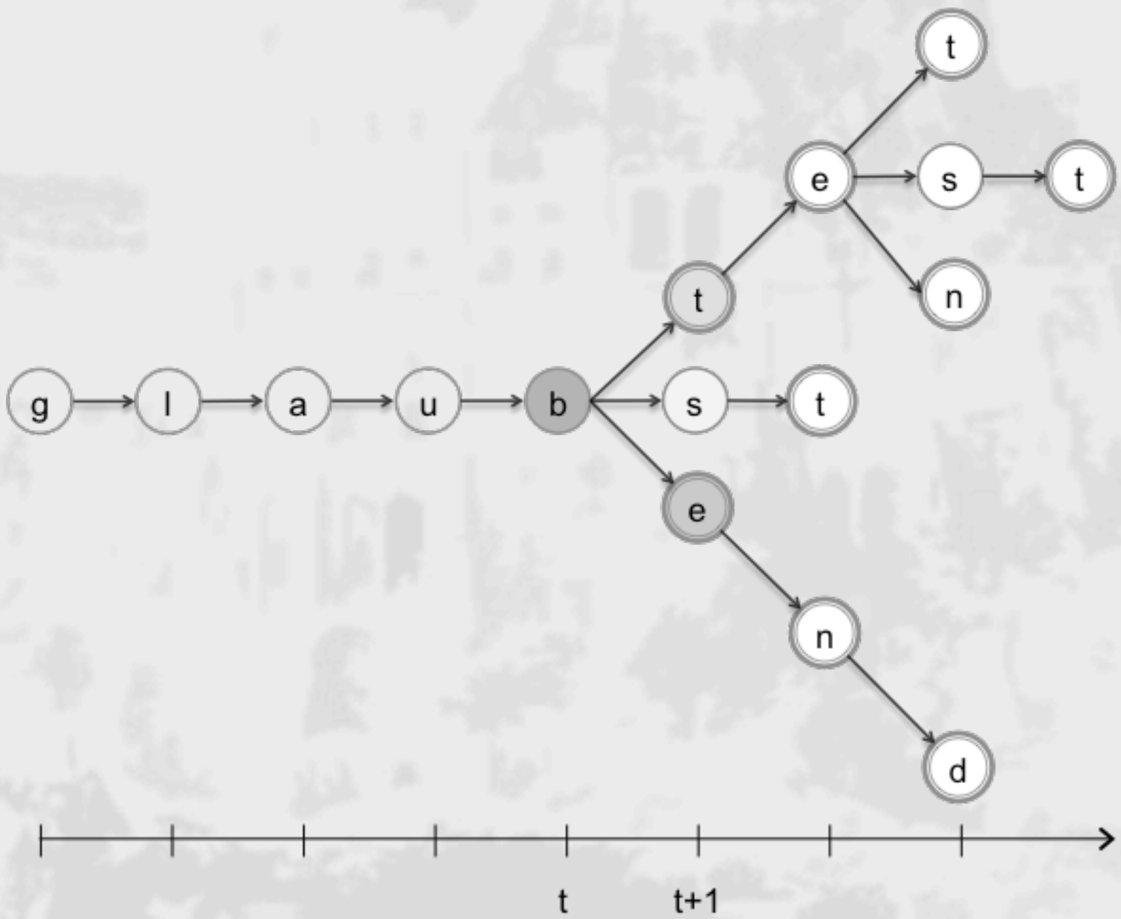


forward connection entropy at morpheme boundary

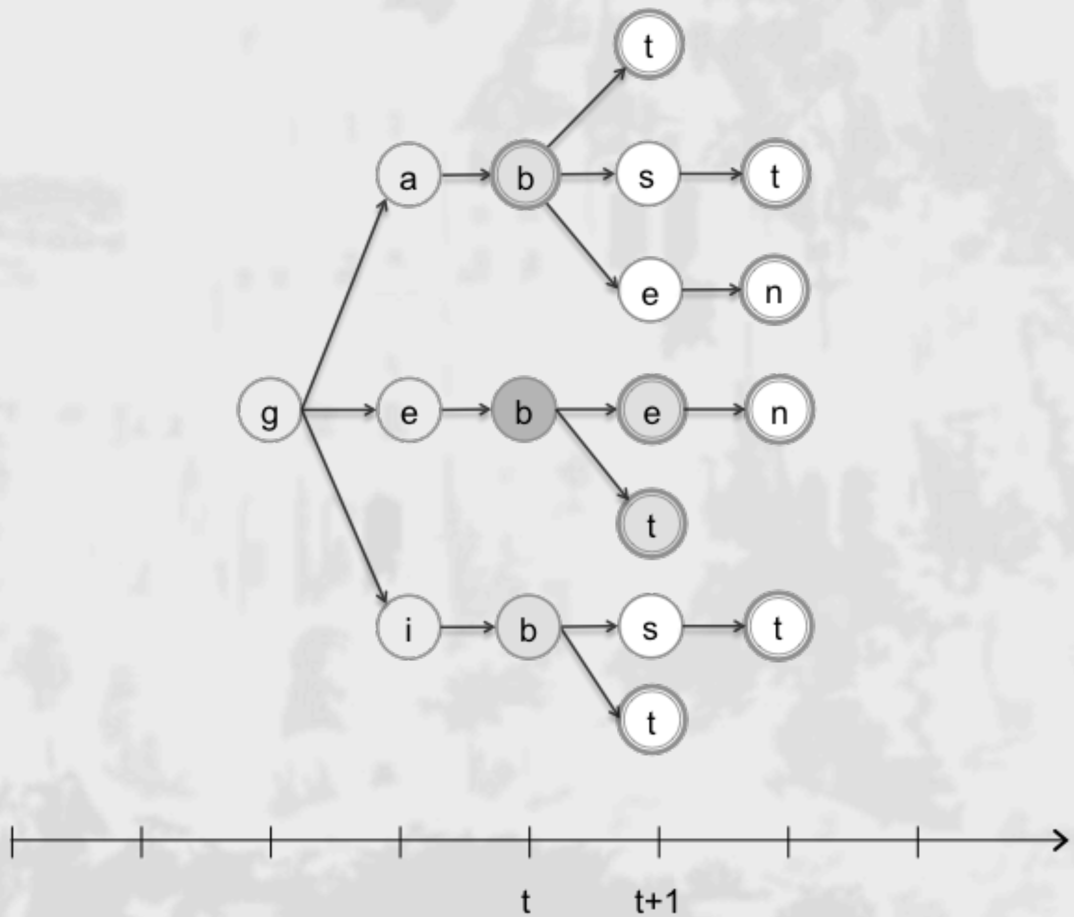


neighbours, serial lexical access and word recall

time & structure (regular)

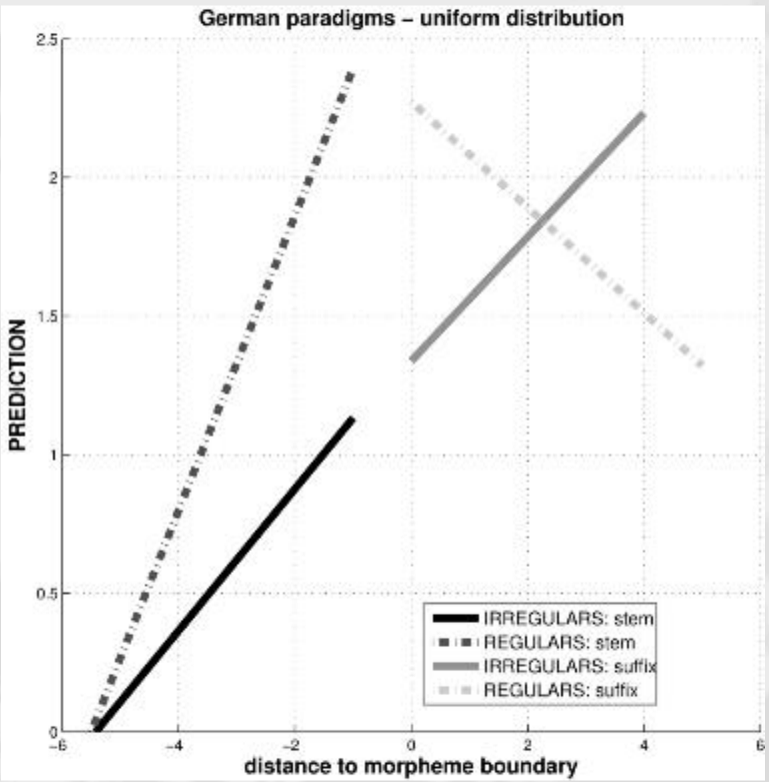


time & structure (irregular)

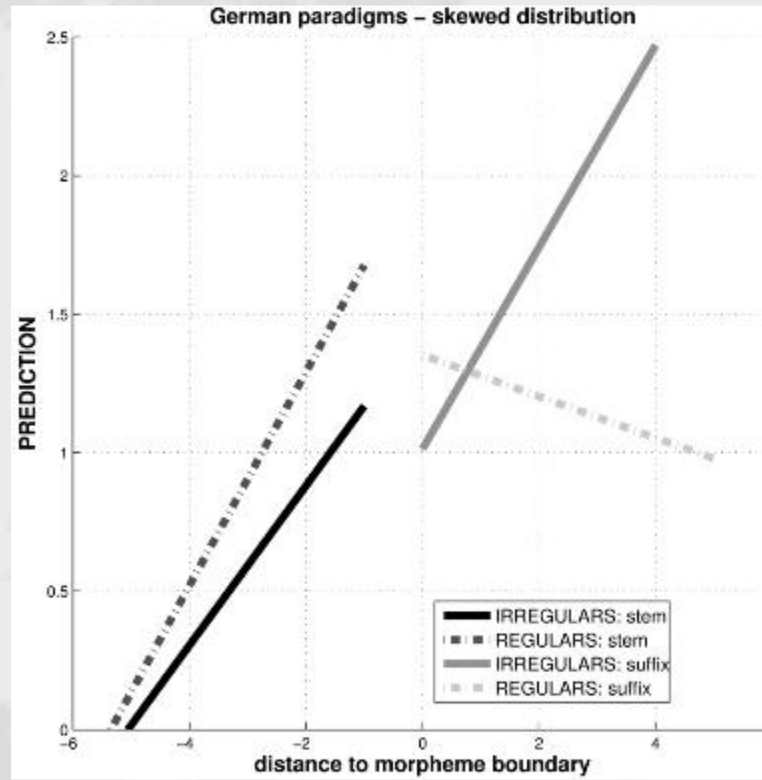


how fast is word completion?

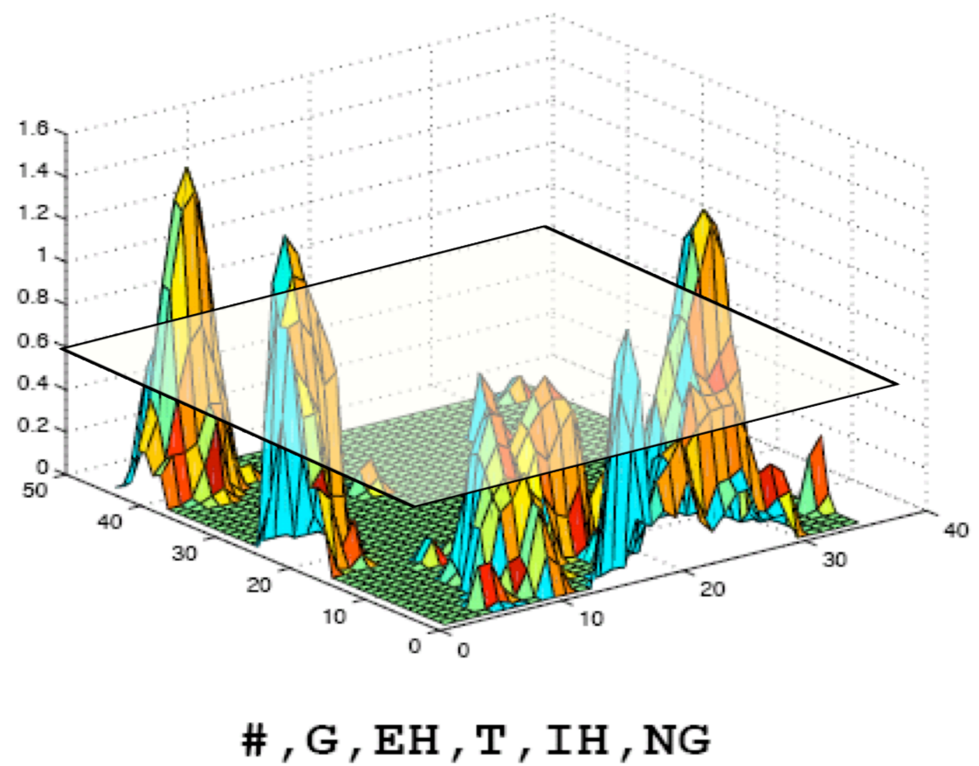
stem-suffix speed-up (UD)



stem-suffix speed-up (SD)

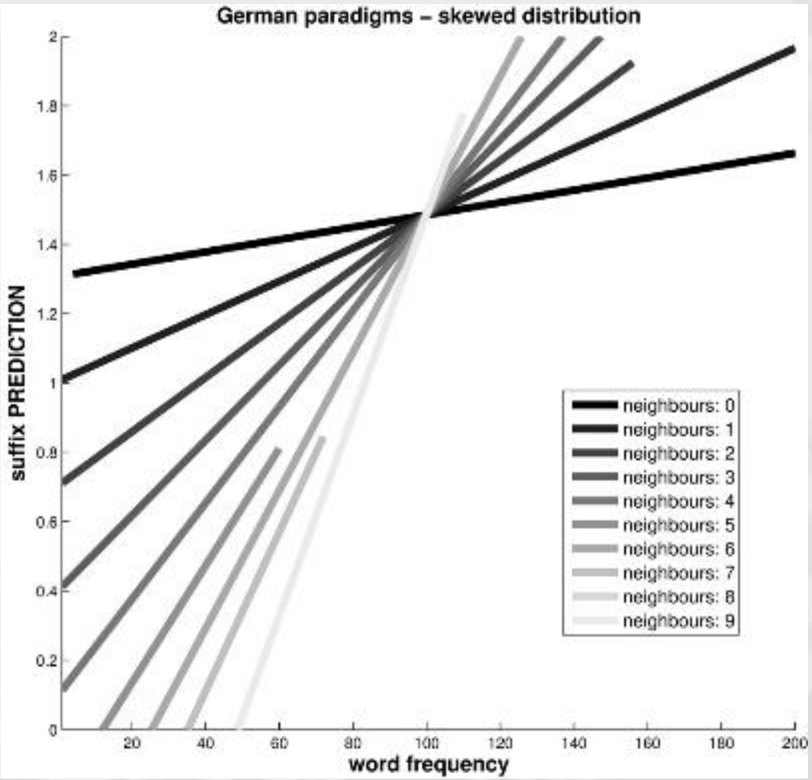


do neighbours always help word filtering?

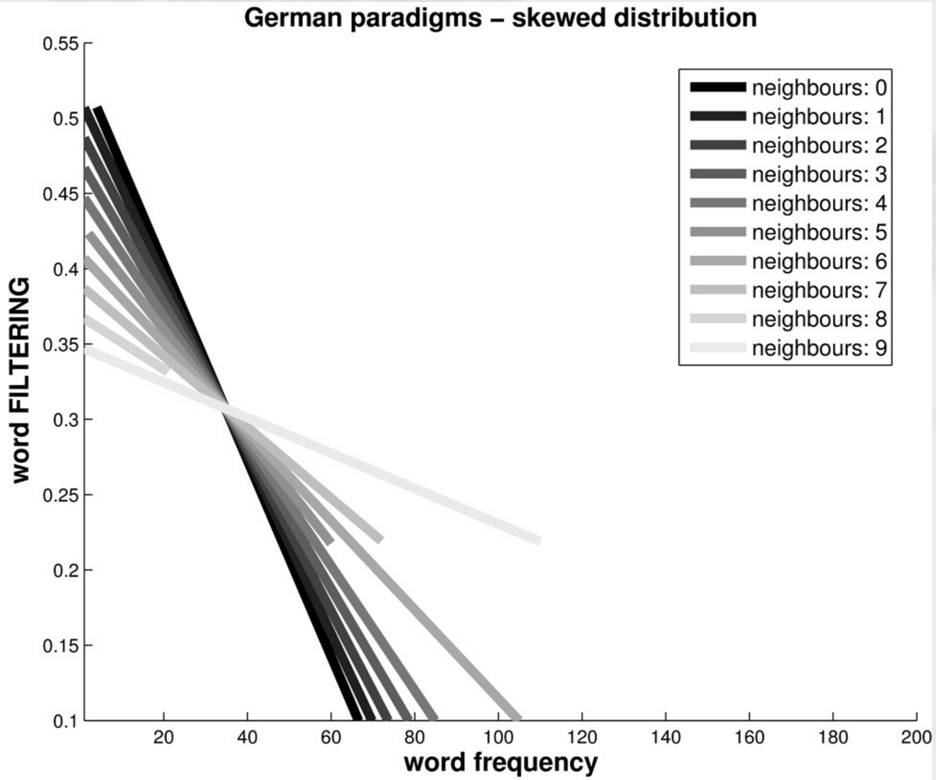


how fast is word completion/recall?

word completion



word recall

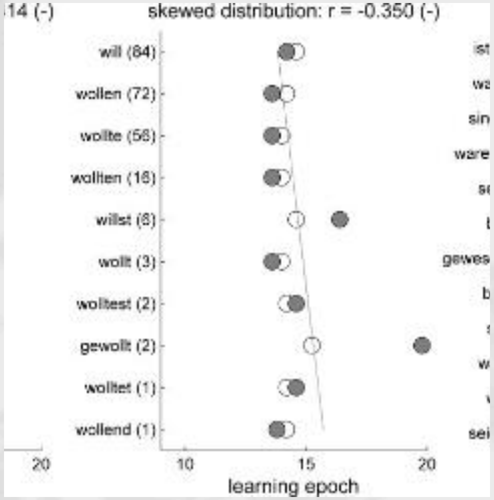


conclusions

words & paradigms

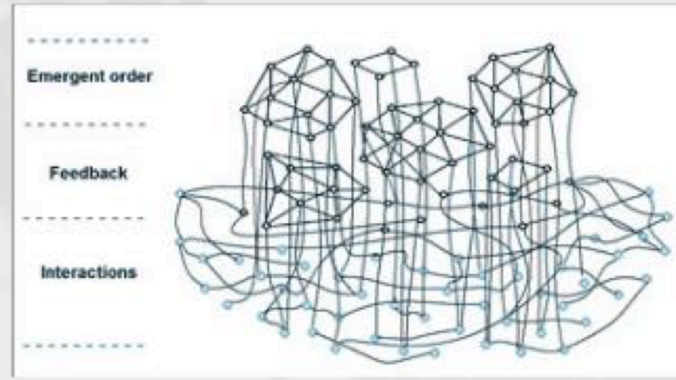
- considerable evidence has accrued on the role of **word paradigms** as both theoretical and cognitive structures regimenting the way words are processed and acquired

- the evidence supports a view of the lexicon as an **emergent integrative system**, where word forms are **concurrently**, **redundantly** and **competitively** stored



integrative memories as lexical maps

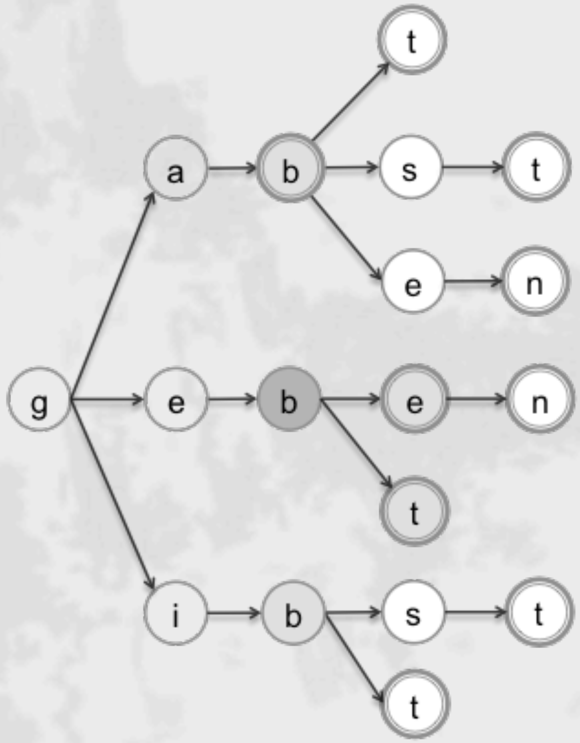
- modelled as integrative memories, lexical maps develop **structure-sensitive self-organisation**
  - no processing vs. memory divide
  - high-frequency words processed through highly-ruotinized firing chains
  - low-frequency words associated with weaker and more “blended” chains



frequency-by-regularity interaction

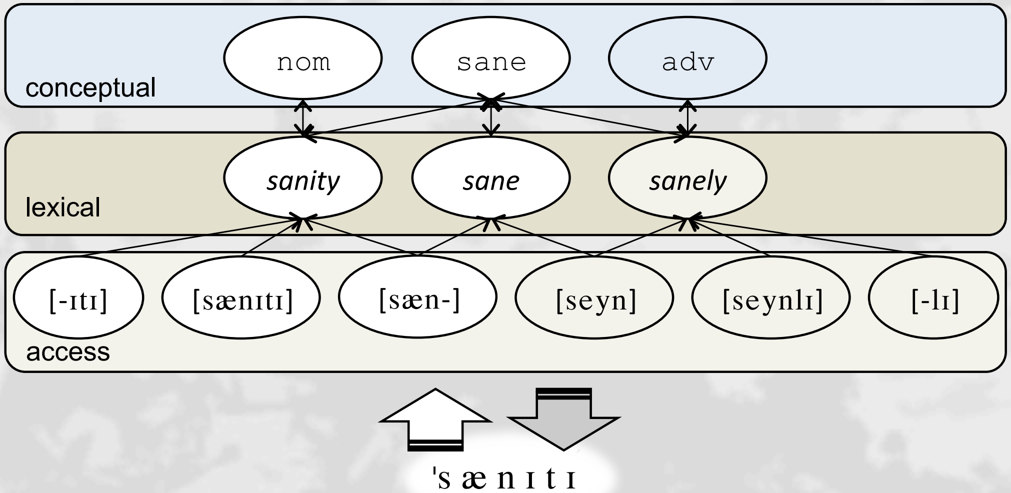
- irregulars
  - processed/stored by **specialised** chains taking advantage of **token frequency** only (no or little shared structure)
  - easier to predict (no branching chains)
  - easier to acquire in the high-frequency range
  - harder to acquire in the low-frequency range

- regulars
  - processed/stored by **blended** chains (shared structure), taking advantage of **type frequency** (family size) and **cumulative family frequency**
  - harder to predict (due to branching effects)
  - easier to acquire (with type frequency additively magnifying effects of token frequency)
  - structure-sensitive effects modulated by family entropy



implications for lexical modelling

- no need to postulate **different underlying mechanisms** for storage/ processing of time-series
  - specific structures are the by-product of **input-specific constraints**
- no need to postulate a principled separation of **regular** (or rule-based) vs. **irregular** (or memory-based) items
  - different behaviour is the outcome of different self-organising patterns of connectivity
- no need to postulate distinct principles of **lexical access/organisation**
  - pre-lexical and post-lexical principles are in fact amenable to underlying unity



joint work with

claudia marzi, marcello ferro,
franco alberto cardillo
and ...

fabian chersi, christina gagné, emmanuel keuleers,
petar milin, giovanni pezzulo, thomas spalding