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

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

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?

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
Speedup is faster for monomorphemic words than for compounds.
speedup is **not** faster for monomorphemic words than for pseudos
influence of bigram frequency

boundary effect
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)
  • observe 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 probabilistic letter distributions significantly affect expectations and surprise for upcoming symbols in the map

• at the word level, compound families make structural discontinuities 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?


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)

\[ \hat{y}_i = \max_{t=1, \ldots, k} \{y_i(t)\} \quad i = 1, \ldots, N \]

symbol level

word level

input layer connections
map nodes connections
IAP connections

\[ X \]
\[ x_1 \]
\[ x_2 \]
\[ \ldots \]
\[ x_j \]
\[ \ldots \]
\[ x_D \]

\[ \# \]
\[ p \]
\[ \ldots \]
\[ o \]
\[ \ldots \]
\[ $ \]

\[ w_1 \]
\[ w_2 \]
\[ \ldots \]
\[ pop \]
\[ \ldots \]
\[ w_L \]
word recall from IAP
general picture

**regulars**

<table>
<thead>
<tr>
<th>Paradigm Acquisition Rank (REGULARS)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>arbeiten (46)</td>
<td></td>
</tr>
<tr>
<td>zeigen (85)</td>
<td></td>
</tr>
<tr>
<td>erreichen (45)</td>
<td></td>
</tr>
<tr>
<td>setzen (66)</td>
<td></td>
</tr>
<tr>
<td>stellen (110)</td>
<td></td>
</tr>
<tr>
<td>versuchen (56)</td>
<td></td>
</tr>
<tr>
<td>glauben (77)</td>
<td></td>
</tr>
<tr>
<td>brauchen (53)</td>
<td></td>
</tr>
<tr>
<td>spielen (62)</td>
<td></td>
</tr>
<tr>
<td>suchen (79)</td>
<td></td>
</tr>
<tr>
<td>führen (71)</td>
<td></td>
</tr>
<tr>
<td>sollen (259)</td>
<td></td>
</tr>
<tr>
<td>hören (66)</td>
<td></td>
</tr>
<tr>
<td>machen (197)</td>
<td></td>
</tr>
<tr>
<td>fragen (72)</td>
<td></td>
</tr>
<tr>
<td>sagen (290)</td>
<td></td>
</tr>
</tbody>
</table>

**German**

<table>
<thead>
<tr>
<th>Paradigm Acquisition Rank (IRREGULARS)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>beginnen (68)</td>
<td></td>
</tr>
<tr>
<td>nennen (58)</td>
<td></td>
</tr>
<tr>
<td>scheinen (61)</td>
<td></td>
</tr>
<tr>
<td>bitten (45)</td>
<td></td>
</tr>
<tr>
<td>heißen (56)</td>
<td></td>
</tr>
<tr>
<td>tun (91)</td>
<td></td>
</tr>
<tr>
<td>mögen (43)</td>
<td></td>
</tr>
<tr>
<td>denken (63)</td>
<td></td>
</tr>
<tr>
<td>kommen (231)</td>
<td></td>
</tr>
<tr>
<td>dürfen (88)</td>
<td></td>
</tr>
<tr>
<td>halten (104)</td>
<td></td>
</tr>
<tr>
<td>bleiben (121)</td>
<td></td>
</tr>
<tr>
<td>tragen (58)</td>
<td></td>
</tr>
<tr>
<td>bestehen (63)</td>
<td></td>
</tr>
<tr>
<td>wissen (109)</td>
<td></td>
</tr>
<tr>
<td>lassen (135)</td>
<td></td>
</tr>
<tr>
<td>schreiben (56)</td>
<td></td>
</tr>
<tr>
<td>nehmen (122)</td>
<td></td>
</tr>
<tr>
<td>fahren (58)</td>
<td></td>
</tr>
<tr>
<td>müssen (289)</td>
<td></td>
</tr>
<tr>
<td>sprechen (95)</td>
<td></td>
</tr>
<tr>
<td>erhalten (62)</td>
<td></td>
</tr>
<tr>
<td>liegen (114)</td>
<td></td>
</tr>
<tr>
<td>geben (243)</td>
<td></td>
</tr>
<tr>
<td>finden (122)</td>
<td></td>
</tr>
<tr>
<td>bringen (99)</td>
<td></td>
</tr>
<tr>
<td>wollen (243)</td>
<td></td>
</tr>
<tr>
<td>stehen (163)</td>
<td></td>
</tr>
<tr>
<td>gehen (177)</td>
<td></td>
</tr>
<tr>
<td>sehen (169)</td>
<td></td>
</tr>
<tr>
<td>können (490)</td>
<td></td>
</tr>
<tr>
<td>haben (1271)</td>
<td></td>
</tr>
<tr>
<td>werden (1344)</td>
<td></td>
</tr>
<tr>
<td>sein (2141)</td>
<td></td>
</tr>
</tbody>
</table>

**irregulars**

<table>
<thead>
<tr>
<th>Paradigm Acquisition Rank (IRREGULARS)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>werden (1344)</td>
<td></td>
</tr>
<tr>
<td>sein (2141)</td>
<td></td>
</tr>
</tbody>
</table>

Trieste 7-15 July 2016
3 paradigms

- **brauchen**
  - Skewed distribution: $r = -0.414$ (-)
  - Even distribution

- **wollen**
  - Skewed distribution: $r = -0.350$ (-)
  - Real distribution

- **sein**
  - Skewed distribution: $r = -0.686$ (*)
  - Real distribution
structure sharing
blending

a) 

b) 

c) 

#,A,B,C,$
#,B,C,$
#,B,$

#,#A,B,$

#,#A,B,$

#,#A,B,$
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

German paradigms – skewed distribution

German paradigms – skewed distribution

word frequency

word frequency

prefix PREDICTION

word FILTERING

neighbours: 0
neighbours: 1
neighbours: 2
neighbours: 3
neighbours: 4
neighbours: 5
neighbours: 6
neighbours: 7
neighbours: 8
neighbours: 9

Trieste 7-15 July 2016
TEX2016
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-routineized 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