Understanding compound words

A new perspective from compositional systems in distributional semantics

Marco Marelli

University of Milano-Bicocca

Compositionality in action

buttercup crown

pineapple pen

Compositionality in action

buttercup crown

pineapple pen

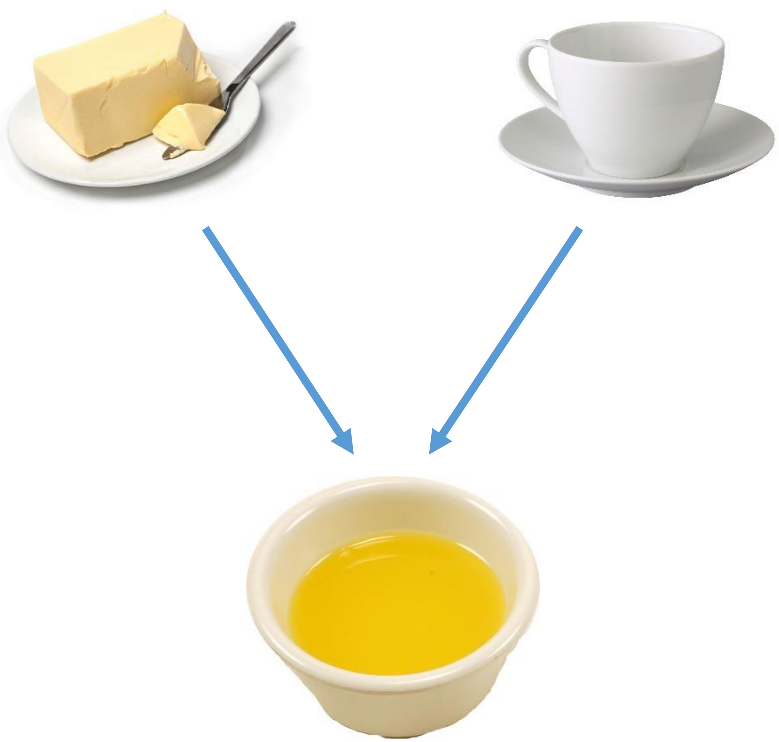
Compositionality in action

buttercup

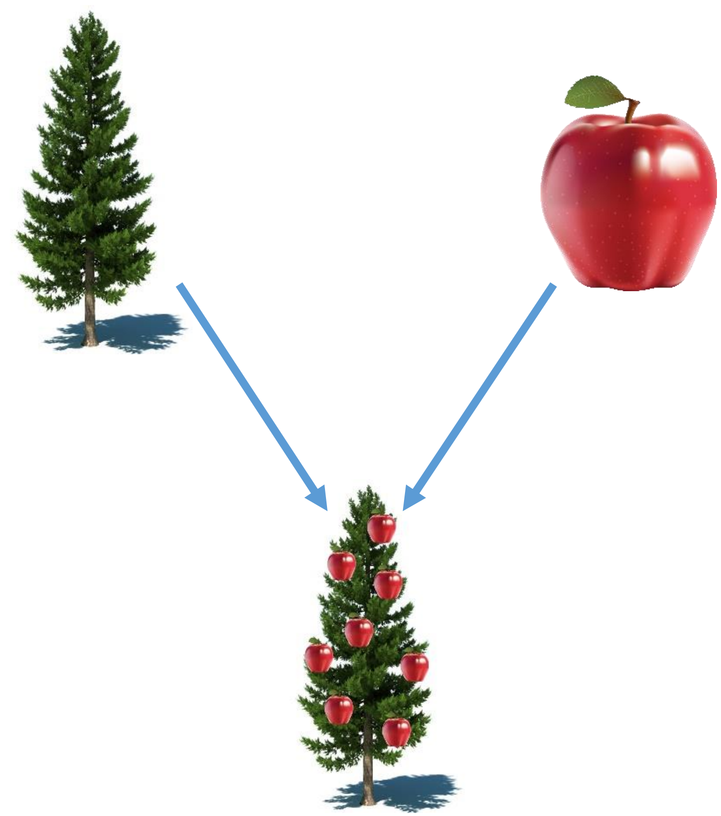
pineapple

Compositionality in action

buttercup



pineapple



Outline

To understand the psycholinguistics of compounding, compositionality is crucial

1. CAOSS: a distributional model to capture internal semantic dynamics in compounds
2. CAOSS simulations of novel compound processing
3. CAOSS-based interpretation of transparency effect on response times and eye-movements in reading

How to model the semantic processing of compounds

(using distributional semantics)

The distributional hypothesis

The meaning of a word is (can be approximated by, learned from) the set of contexts in which it occurs

We found a little, hairy ***wampimuk***

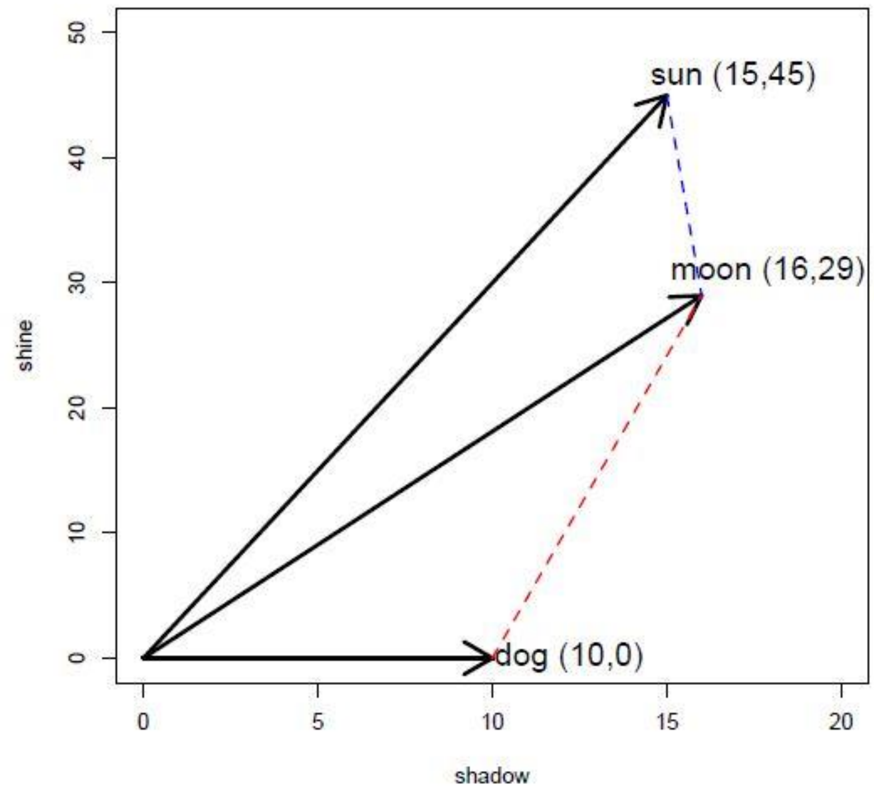
sleeping behind the tree

The foundations of distributional semantics

- The distributional hypothesis can be formalized through computational methods:
  - Word meanings are modelled through lexical cooccurrences
  - In turn, lexical cooccurrences can be collected from linguistic corpora

The geometry of meaning

	shadow	shine
moon	16	29
sun	15	45
dog	10	0



A model of the conceptual system?

- Very appealing for cognitive science
- Plausible nuanced representations for meanings
- Related to biologically plausible learning-mechanism

- Distributional approaches very effective in many cognitive experiments
  - explicit semantic intuitions (Landauer and Dumais, 1997)
  - learning curves (Landauer and Dumais, 1997)
  - fixation times in reading (Griffiths et al., 2007)
  - priming paradigms (Jones et al., 2006)

Distributional semantics for compounding?

- Language is a productive system, but vanilla distributional models cannot induce representations for novel combinations

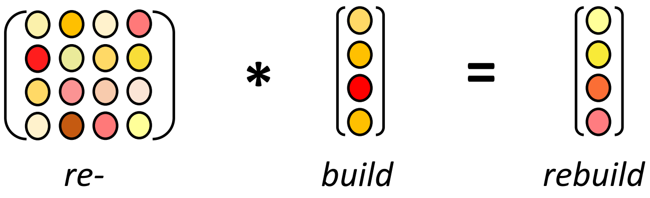
- Lynott & Ramscar (2001): distributional semantics cannot account for effects in compound-processing

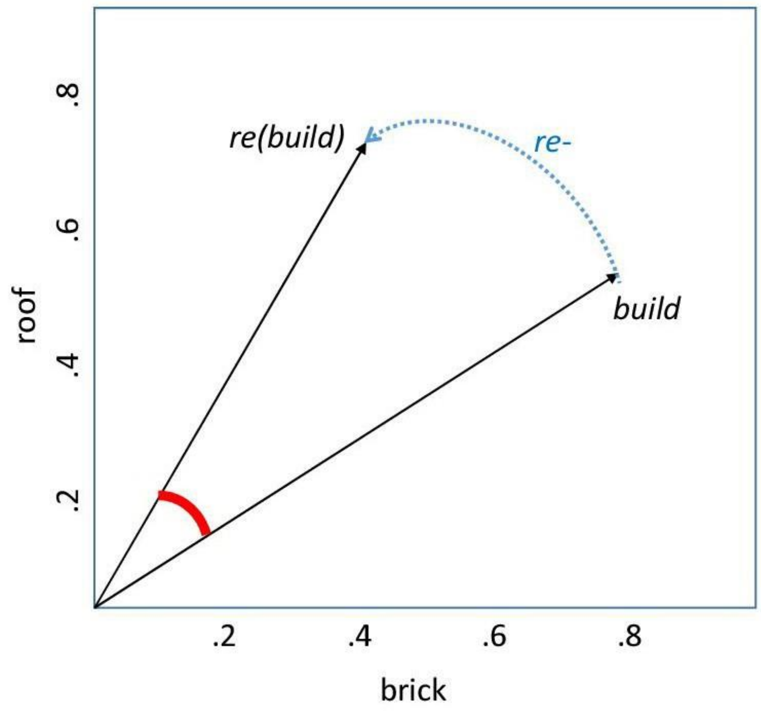
SOLUTION: compositional distributional semantics

Compositional distributional models

- Recently, several proposals in computational linguistics
  - For example, simple sums or multiplication of constituent vectors (Mitchell & Lapata, 2010)
- In psycholinguistics, function-based FRACSS model (Marelli & Baroni, 2015)
  - Account for several morphology effects, including response times and priming effects

The FRACSS model





Why a different approach for compounds?

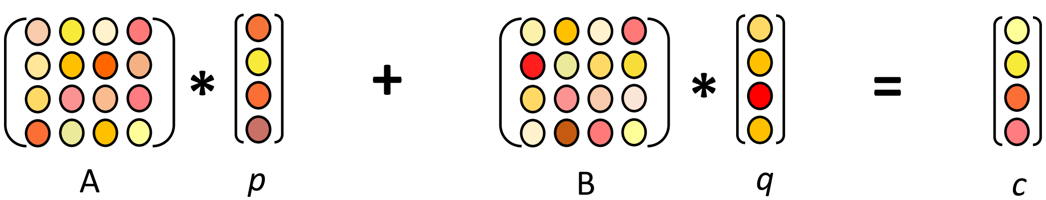
- A model for compound meanings should be able to account for:
  - The productivity of the system
  - The ease of comprehension of novel compounds
  - The possibility to generate compounds including newly acquired words (<u>out of the possibilities of function models</u>)
  - Impact of constituent order (<u>out of the possibilities of simpler proposals</u>)

Function-based and simpler models are not an ideal solution for compounding

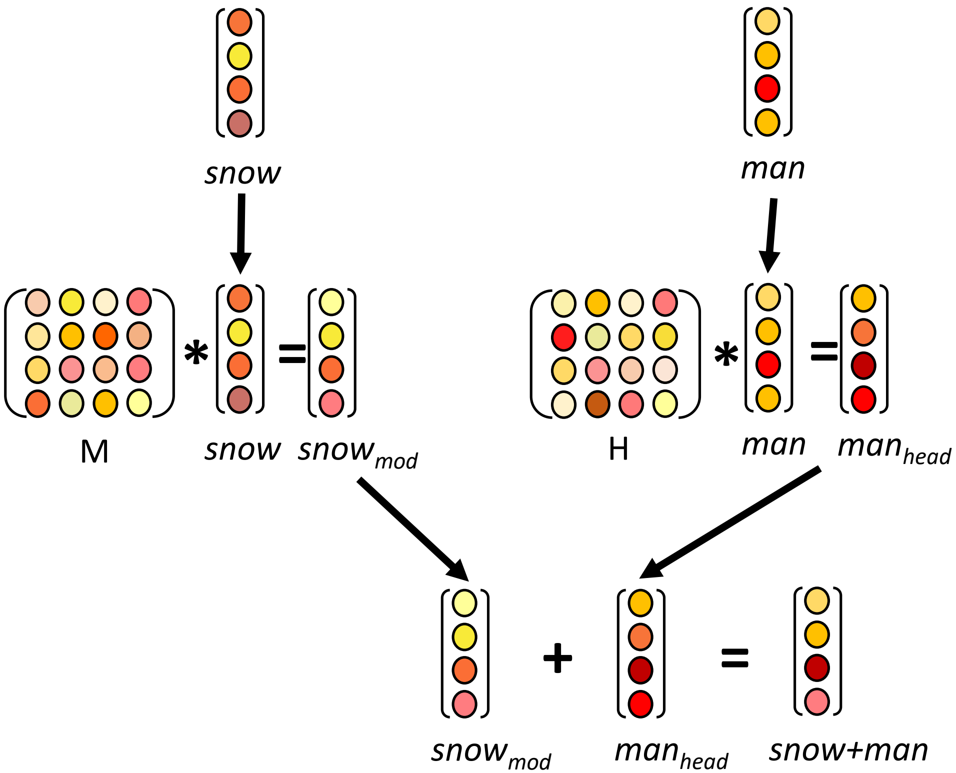
Guevara (2011)

We turn to the system proposed by Guevara (2011)

A compositional representation is obtained through a semantic update of the constituents, achieved by means of a set of weight matrices



CAOSS: Compounding as Abstract Operation in Semantic Space



STEP 0
semantic representations for independent words

STEP 1
role-dependent update by means of CAOSS matrices

STEP 2
combination of the obtained constituent representations

CAOSS training

$$
\begin{array}{ccccccccc}
\text{carwash} \begin{pmatrix} ca_1 \\ ca_2 \\ ca_3 \\ \dots \\ ca_N \end{pmatrix} & & & & \begin{pmatrix} ua_1 \\ ua_2 \\ ua_3 \\ \dots \\ ua_N \end{pmatrix} \text{car} & & & & \begin{pmatrix} va_1 \\ va_2 \\ va_3 \\ \dots \\ va_N \end{pmatrix} \text{wash} \\
\text{swordfish} \begin{pmatrix} cb_1 \\ cb_2 \\ cb_3 \\ \dots \\ cb_N \end{pmatrix} & = & \begin{pmatrix} m_{11} & m_{12} & m_{13} & \dots & m_{1N} \\ m_{21} & \dots & \dots & \dots & \dots \\ m_{31} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ m_{N1} & \dots & \dots & \dots & m_{NN} \end{pmatrix} & * & \begin{pmatrix} ub_1 \\ ub_2 \\ ub_3 \\ \dots \\ ub_N \end{pmatrix} \text{sword} & + & \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1N} \\ h_{21} & \dots & \dots & \dots & \dots \\ h_{31} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ h_{N1} & \dots & \dots & \dots & h_{NN} \end{pmatrix} & * & \begin{pmatrix} vb_1 \\ vb_2 \\ vb_3 \\ \dots \\ vb_N \end{pmatrix} \text{fish} \\
\dots & & M & & \dots & & H & & \dots \\
\text{moonlight} \begin{pmatrix} cz_1 \\ cz_2 \\ cz_3 \\ \dots \\ cz_N \end{pmatrix} & & & & \begin{pmatrix} uz_1 \\ uz_2 \\ uz_3 \\ \dots \\ uz_N \end{pmatrix} \text{moon} & & & & \begin{pmatrix} vz_1 \\ vz_2 \\ vz_3 \\ \dots \\ vz_N \end{pmatrix} \text{light}
\end{array}
$$

CAOSS: a psycholinguistic evaluation

(1) The processing of novel compounds

Novel compounds: roles and relations

Constituent roles

Head (rightmost element):

A *mountaine magazine* is a *magazine*

Modifier (leftmost element):

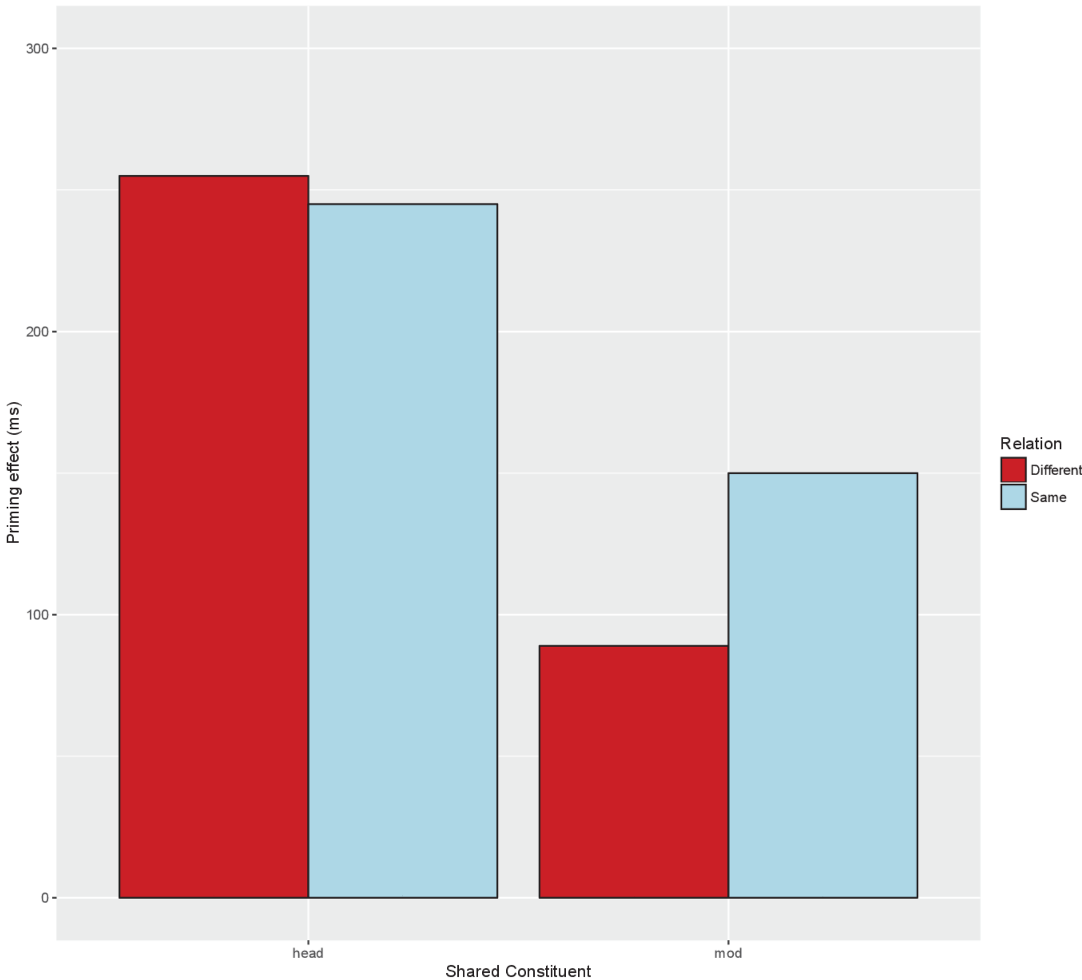
A *mountain magazine* has something to do with *mountains*

Compound relations

Unexpressed links between head and modifier

A *mountain magazine* is a *magazine* about *mountain*

Relational priming effect



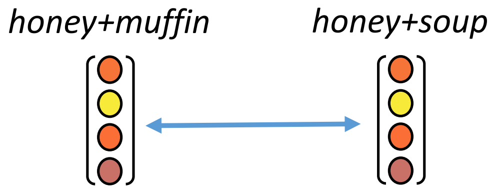
Behavioral results from Gagné (2001)

Primes for the target *honey soup*

Shared Constituent	**Relation**	**Prime Example**
modifier	same	*honey muffin*
modifier	different	*honey insect*
head	same	*ham soup*
head	different	*holiday soup*

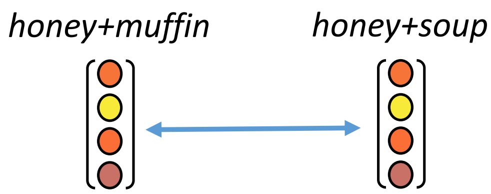
Relational priming effect in CAOSS

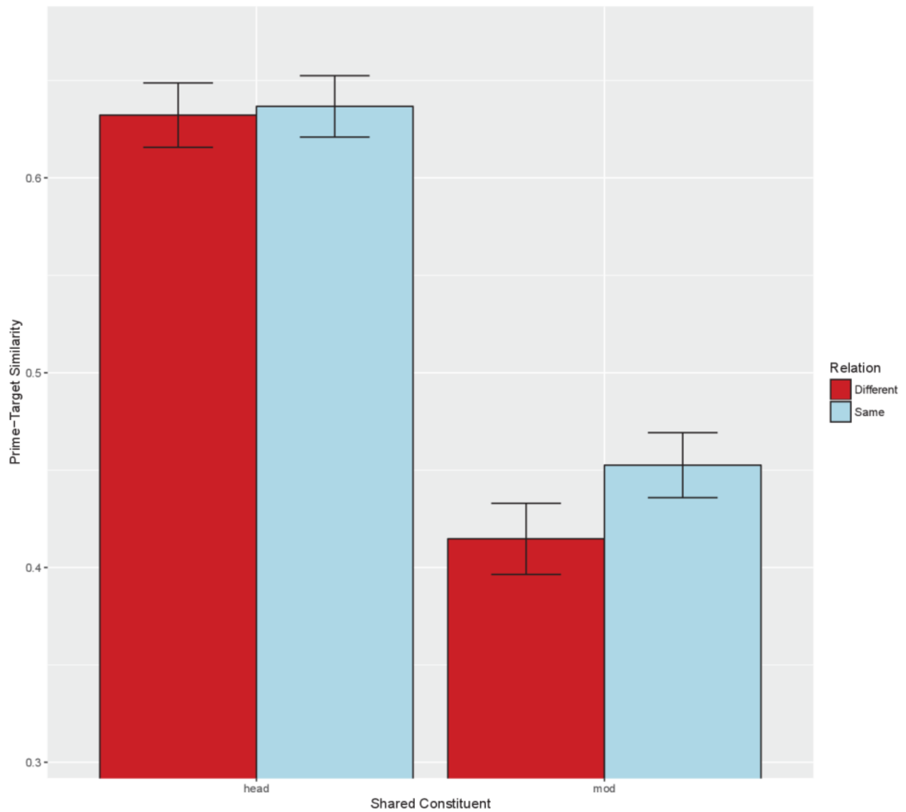
Priming effect as similarity between compositional meanings

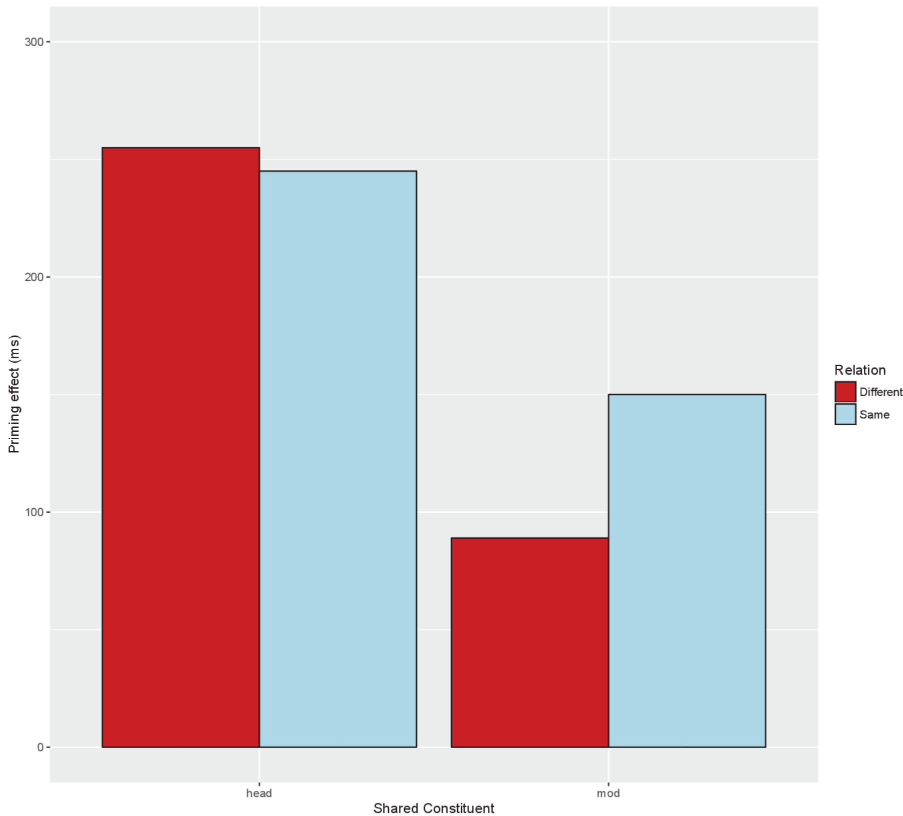


Relational priming effect in CAOSS

Priming effect as similarity between compositional meanings

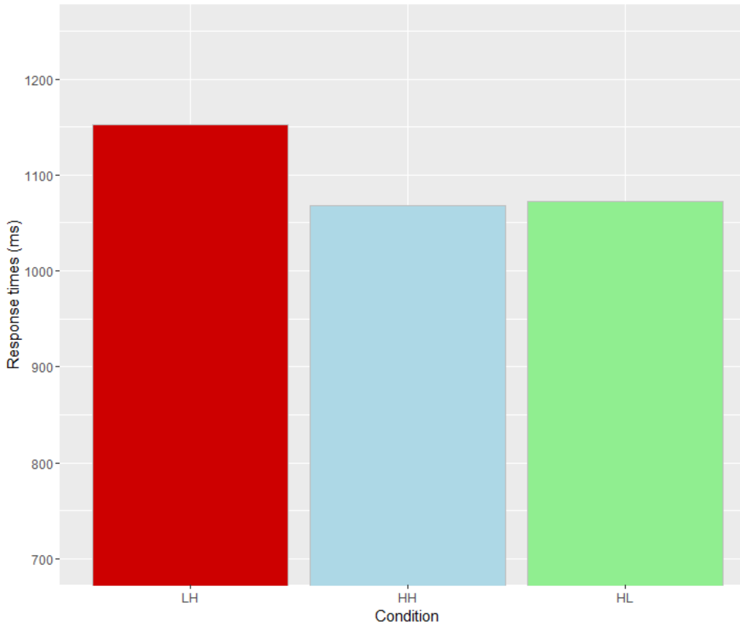






Relational dominance effect

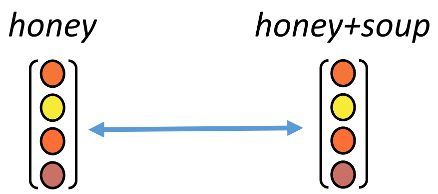
Behavioral results from Gagné & Shoben (1997)



Condition	Target Example	Dominant Relation for Modifier	Dominant Relation for Head	Actual Relation
LH	*plastic crisis*	MADE-OF	ABOUT	ABOUT
HH	*plastic toy*	MADE-OF	MADE-OF	MADE-OF
HL	*plastic equipment*	MADE-OF	FOR	MADE-OF
LH	*college headache*	ABOUT	CAUSED-BY	CAUSED-BY
HH	*college magazine*	ABOUT	ABOUT	ABOUT
HL	*college treatment*	ABOUT	FOR	IN

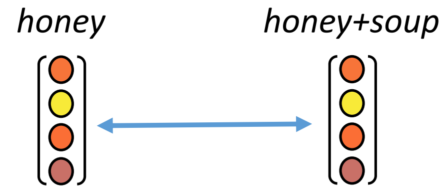
Relational dominance in CAOSS

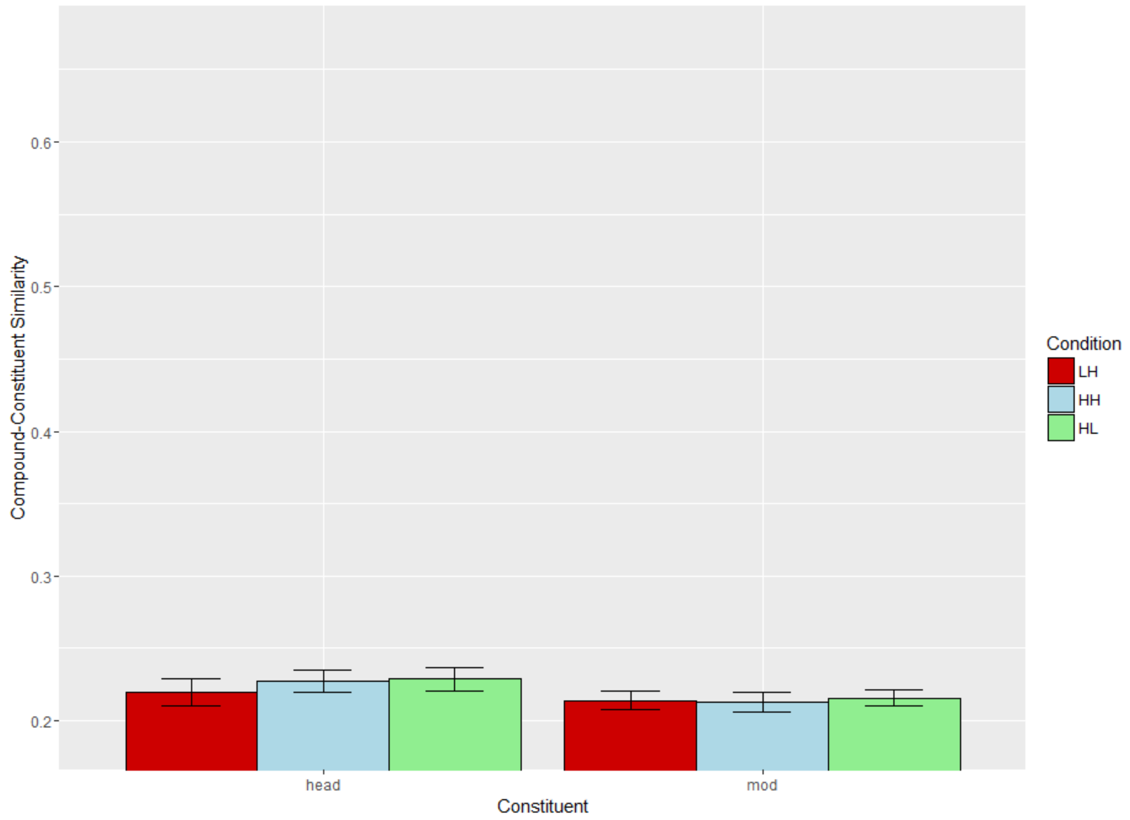
Relational dominance as similarity between constituents and compositional meanings



Relational dominance in CAOSS

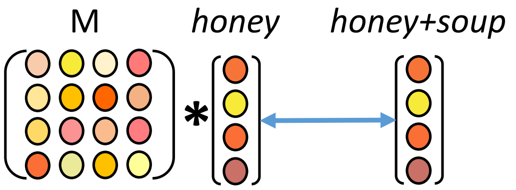
Relational dominance as similarity between constituents and compositional meanings





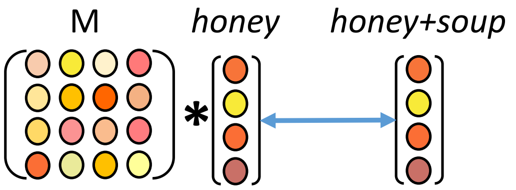
Relational dominance in CAOSS

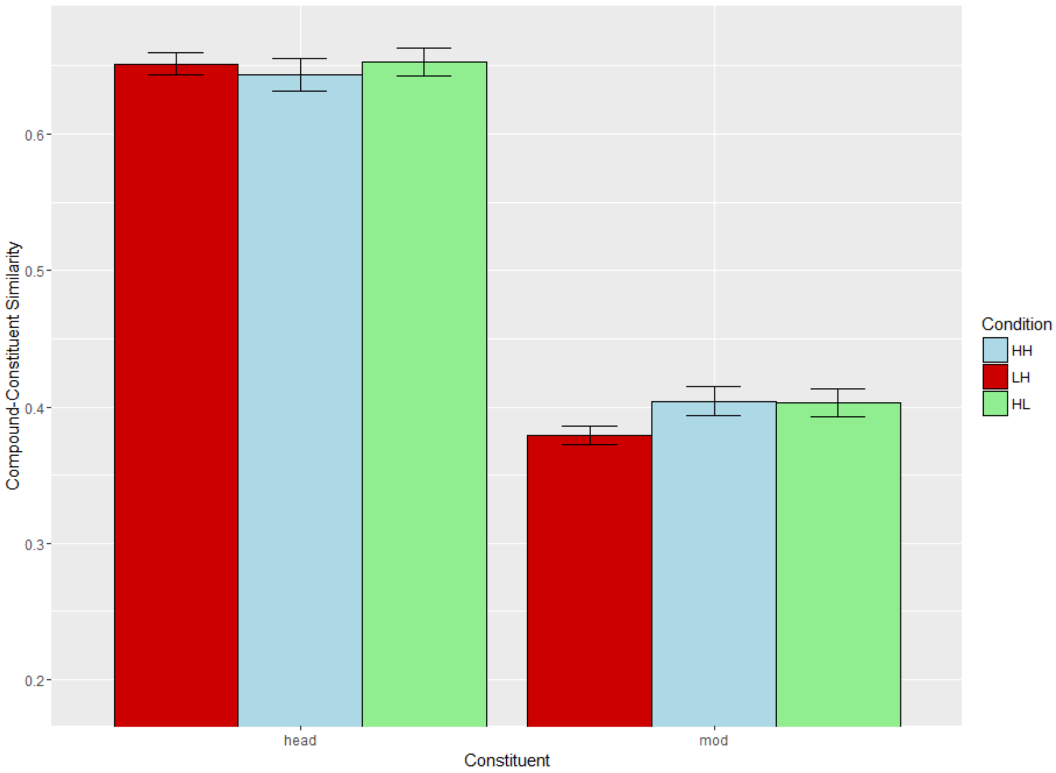
Relational dominance as similarity between updated constituents and compositional meanings

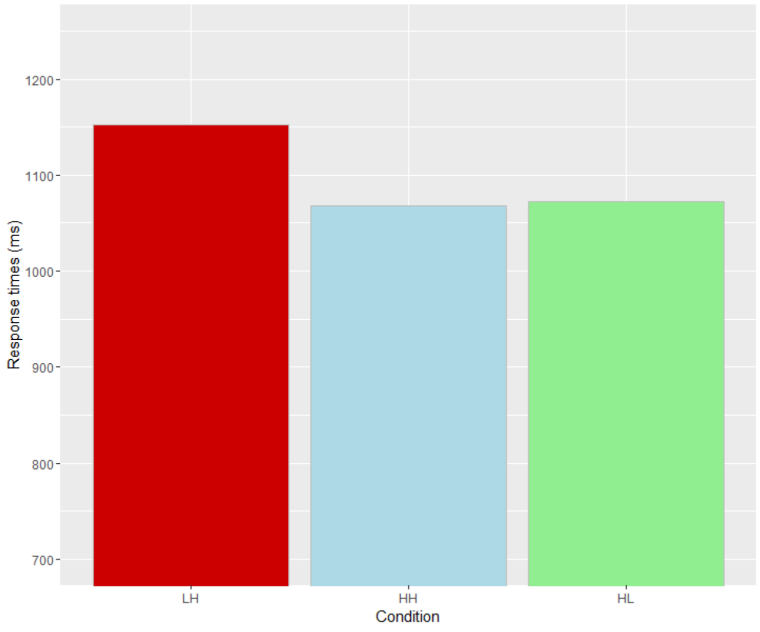


Relational dominance in CAOSS

Relational dominance as similarity between updated constituents and compositional meanings







CAOSS and novel compounds

- CAOSS can provide apt representations for novel combinations in a data-driven framework
- Psycholinguistic effects are mirrored in CAOSS predictions
- Compound relations and head-modifier roles can be seen as by-products of compound usage, or high-level description of a nuanced compositional system

CAOSS: a psycholinguistic evaluation

(2) The processing of familiar compounds

Semantic transparency in chronometric studies

- Evidence of transparency effects is at times inconsistent (e.g., Zwitserlood, 1994; Pollatsek & Hyona 2005)

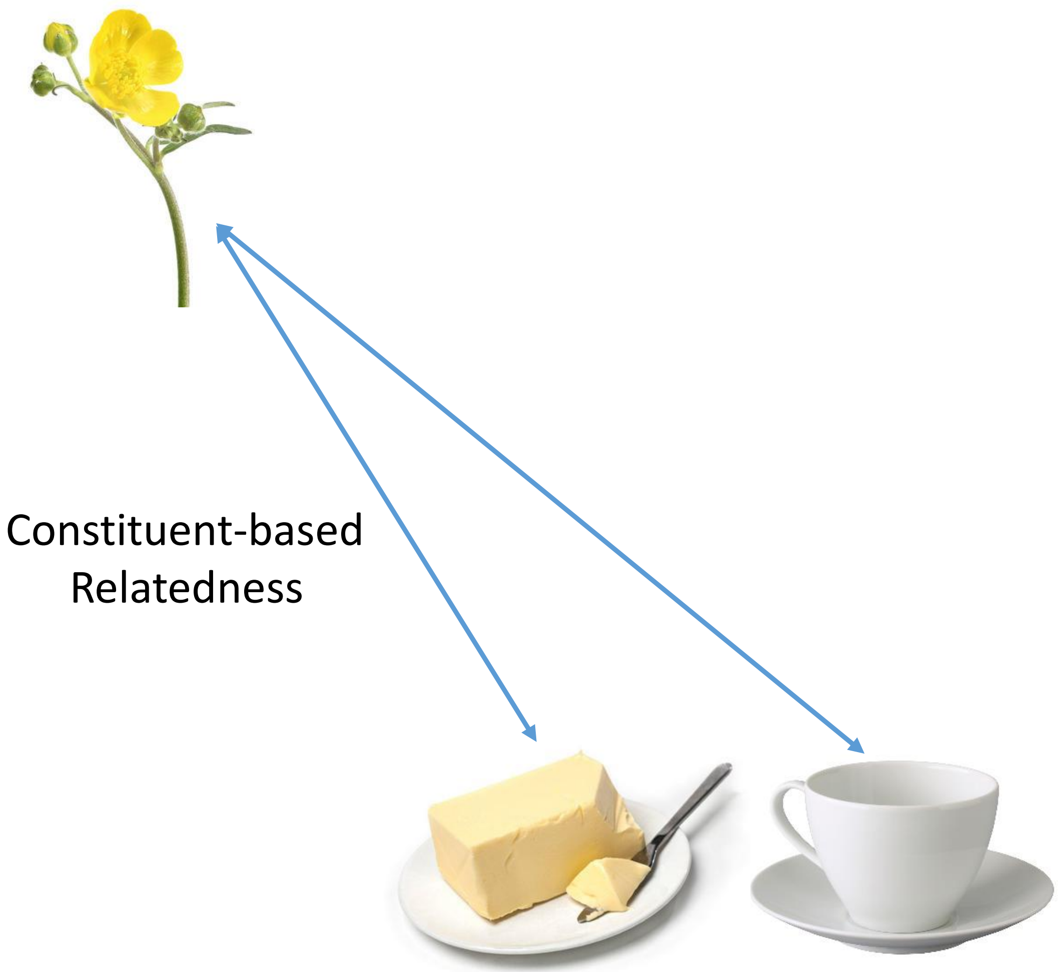
- When an effect is observed, is often characterized in compositional terms by means of:
  - rating instructions (Marelli & Luzzatti, 2012)
  - experimental design (Frisson et al., 2008; Ji et al., 2011)
  - training examples in modelling (Marelli et al., 2014)

Compositionality may play a crucial role in a cognitively-relevant definition of semantic transparency

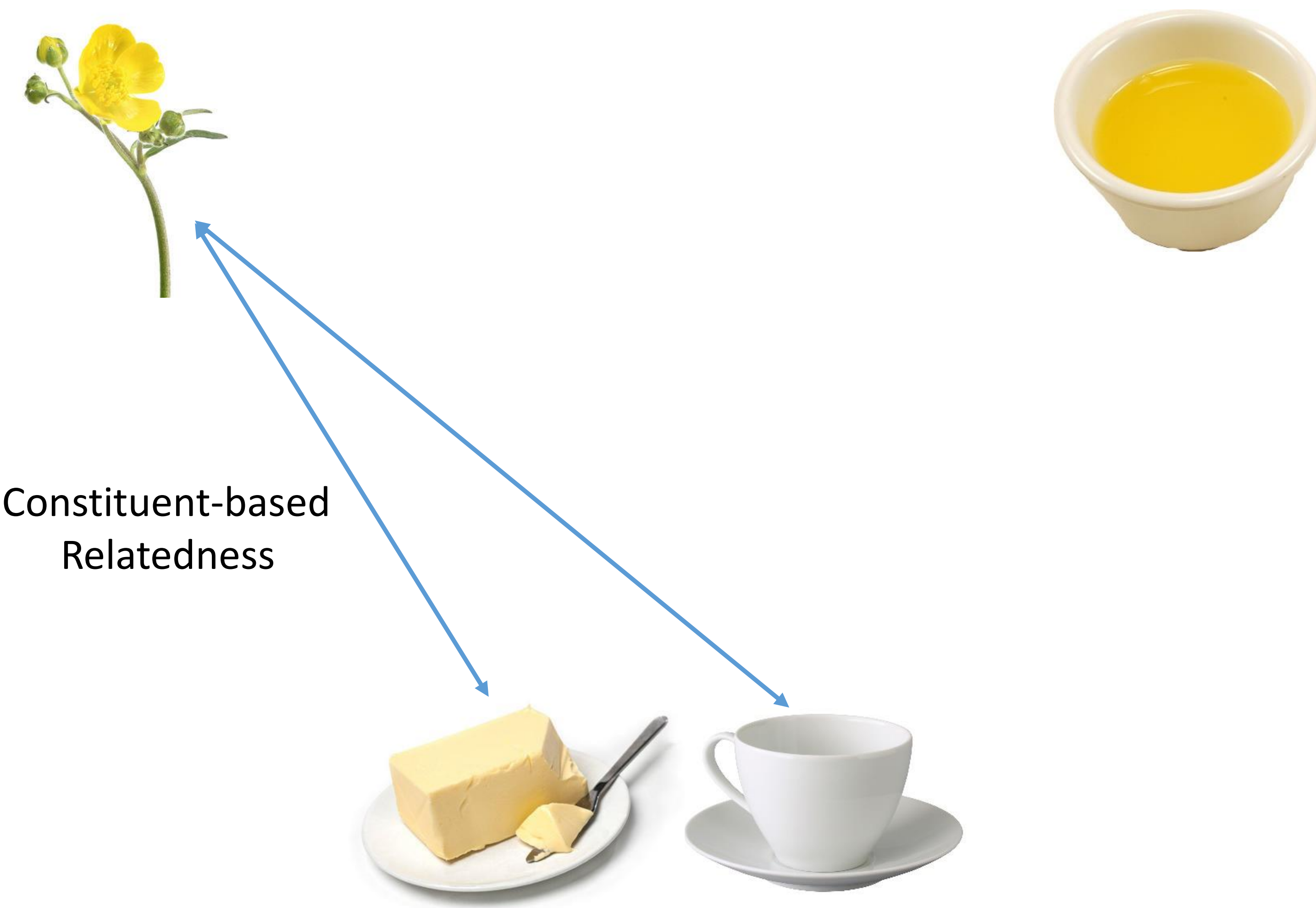
Why compositionality?

- The compositional procedure should be **fast and automatic**: generating new meanings is the very purpose of compounding
- A compositional meaning **should be always computed** by the speaker: when processing a compound, the speaker cannot know in advance whether it is familiar or not
- Such a procedure would be **most often effective**: very opaque compounds are rare, and the meaning of partially opaque words can be approximated compositionally

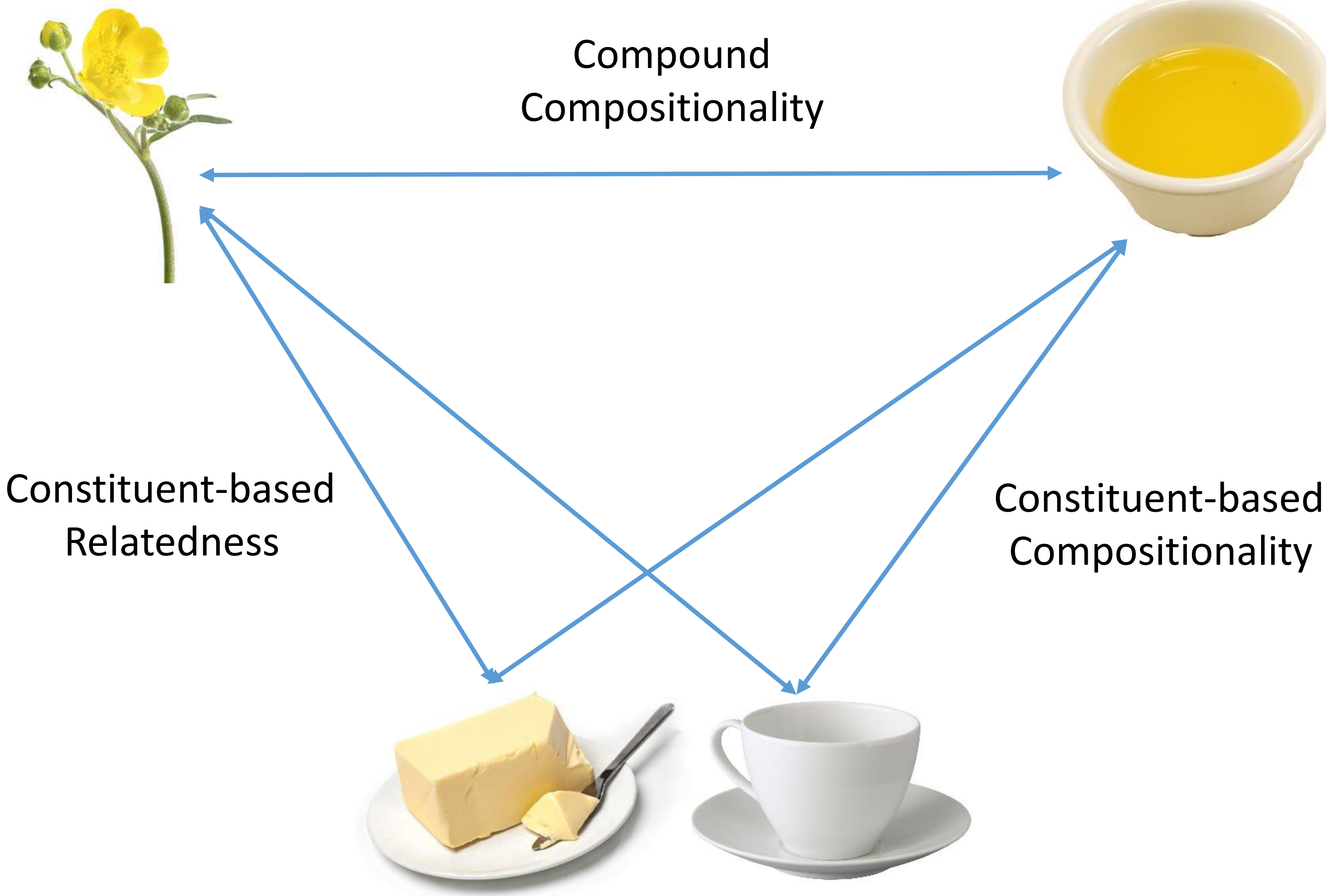
The many faces of transparency



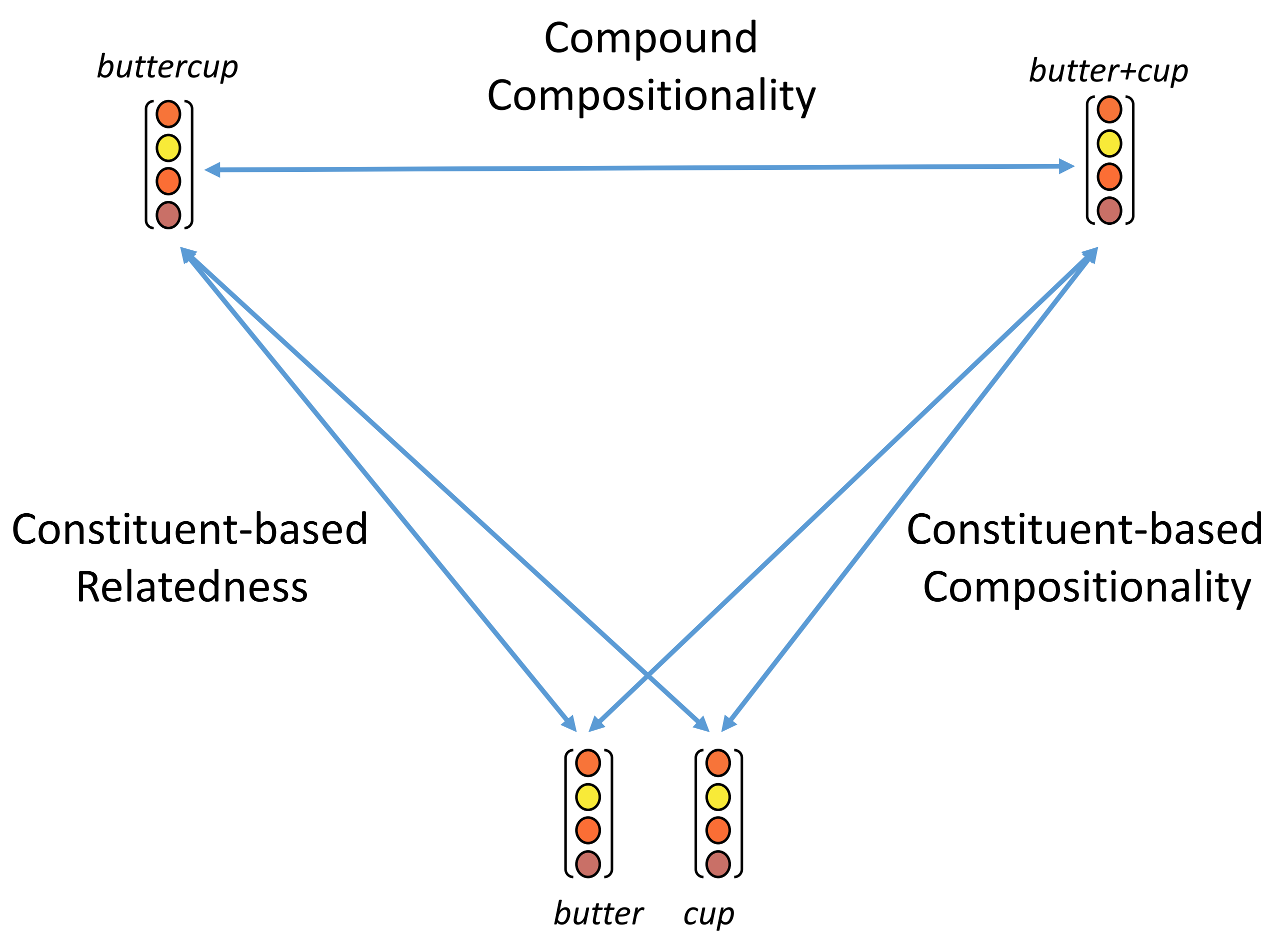
The many faces of transparency



The many faces of transparency

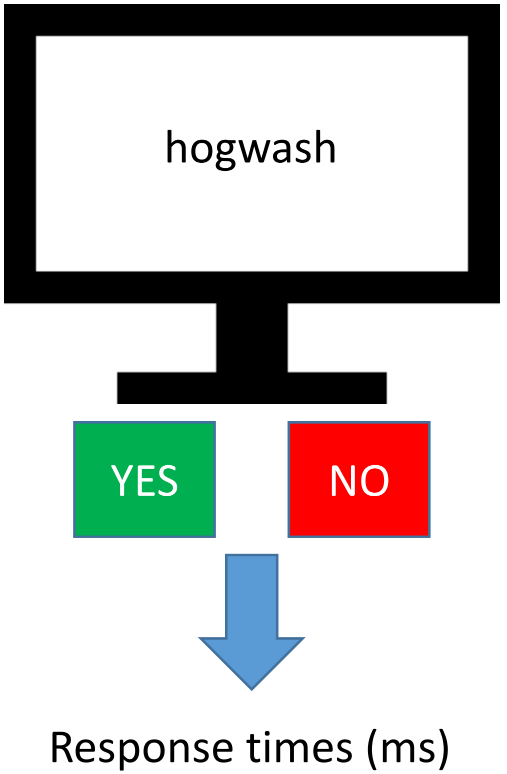


The many faces of transparency in CAOSS

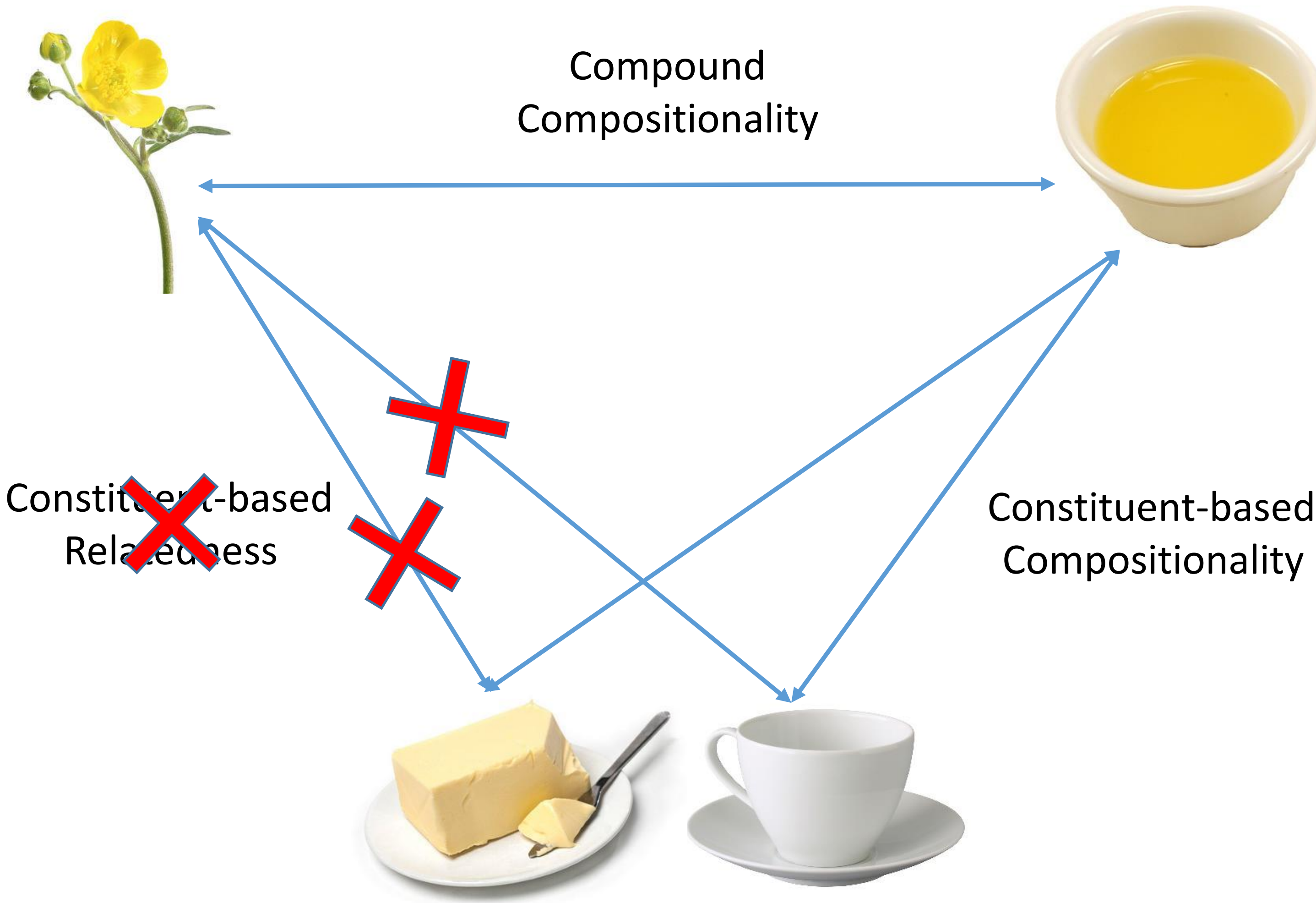


CAOSS and lexical decision

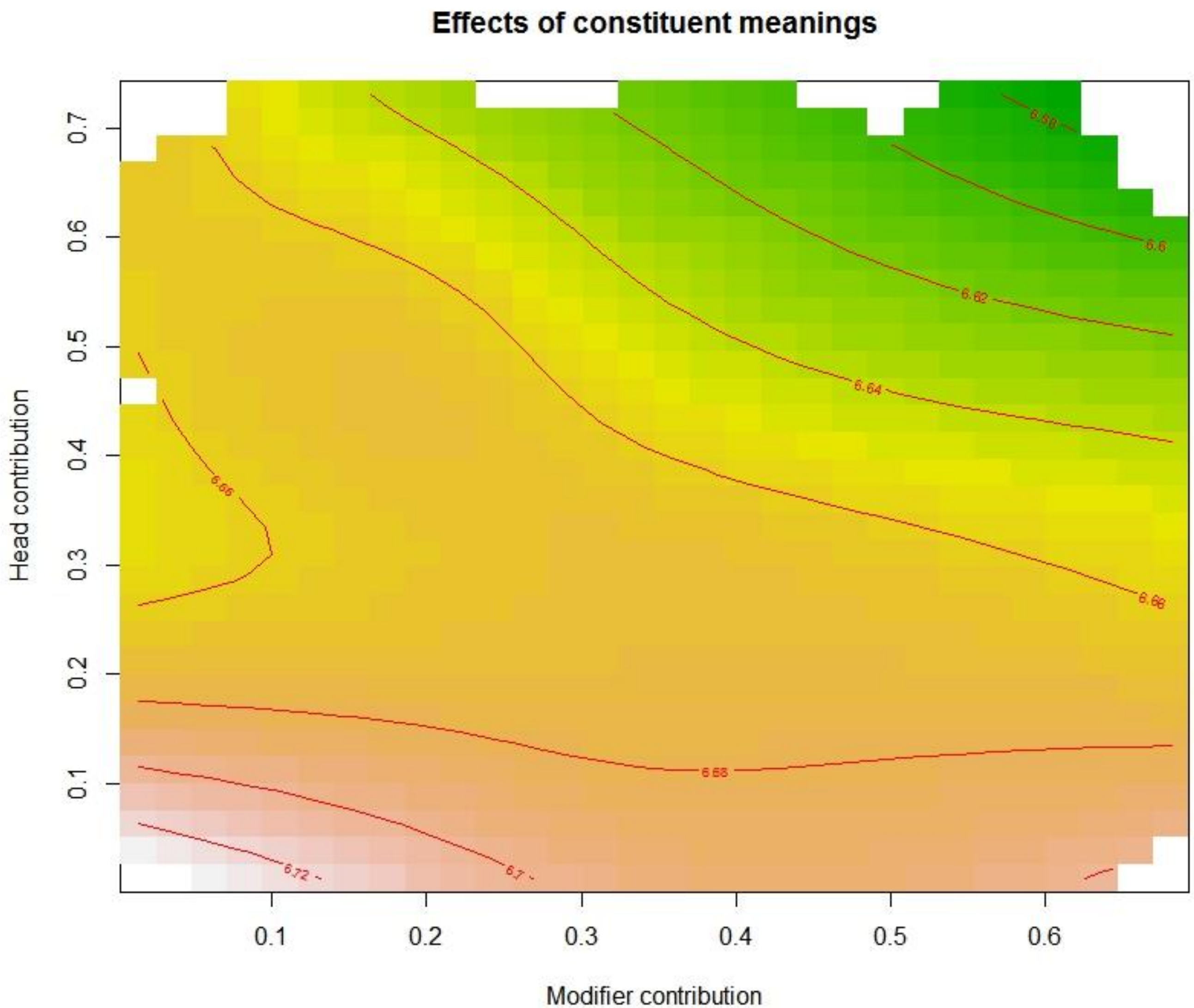
- Response times for 1845 lexicalized compounds from the English Lexicon Project (Balota et al., 2007)
- Semantic effects tested against a baseline of form-related variables (length, frequency, etc)



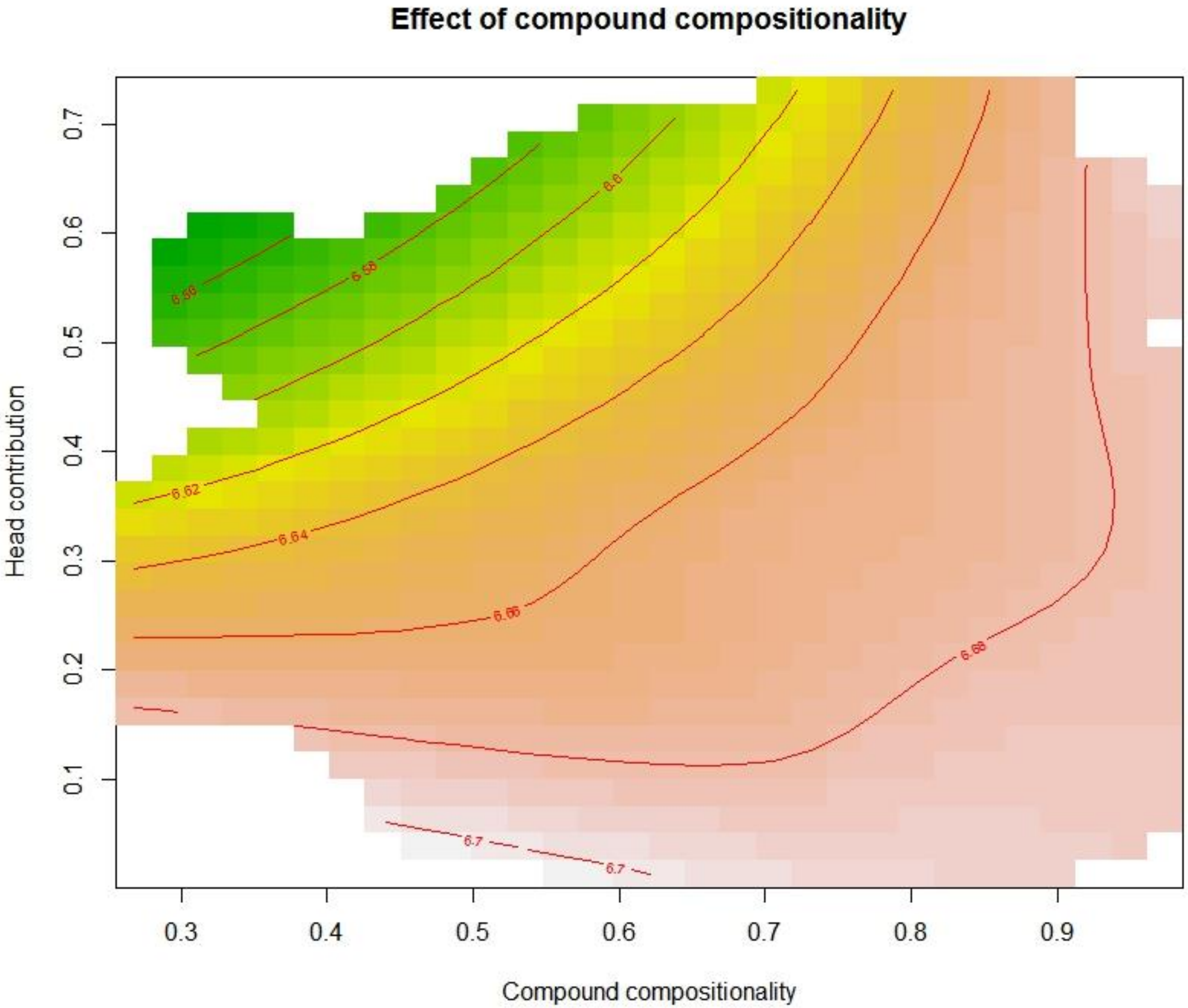
CAOSS effects in lexical decision



CAOSS effects in lexical decision



CAOSS effects in lexical decision



CAOSS effects in lexical decision

- Compound compositionality affects response times
- The constituent impact is better explained in terms of *their contribution to the compositonal meaning*
- Head constituent has a modulating role

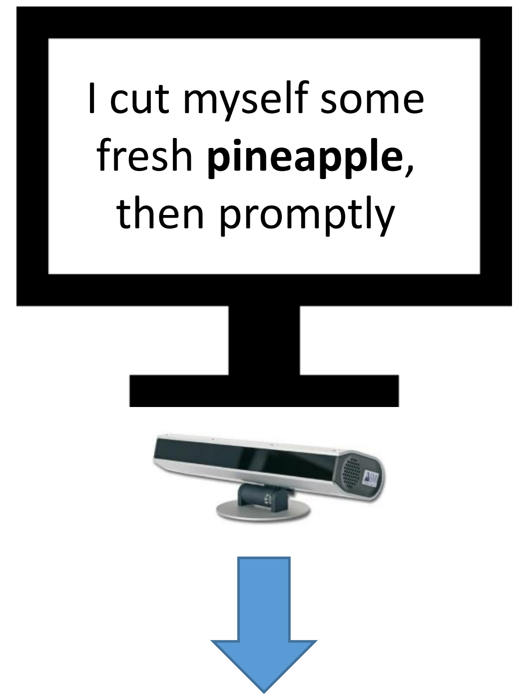
CAOSS effects in lexical decision

- The compositionality effect is unexpected: lack of compositionality eases recognition!
- Task effect?
  - any string activating much semantic information is likely to be a word
  - low compositionality means that a compound activate two different meanings
  - large semantic activation boosts response times

CAOSS and eye tracking

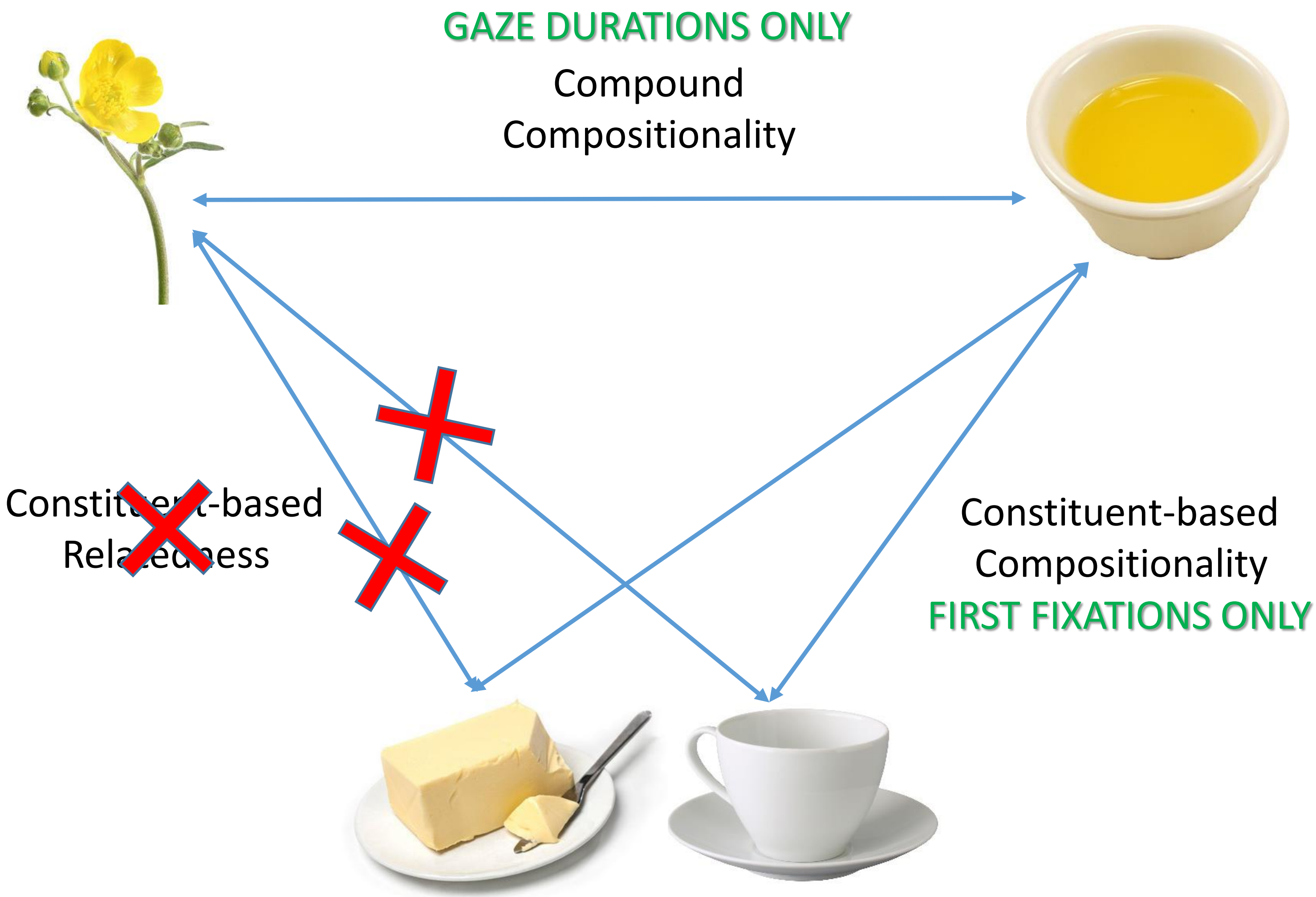
- Response times for 78 lexicalized compounds from GECO (Cop et al., in press)
- Semantic effects tested against a baseline of form-related variables

- Two models:
    - **first fixation times** as index of early processing
    - **gaze durations** as index of late processing

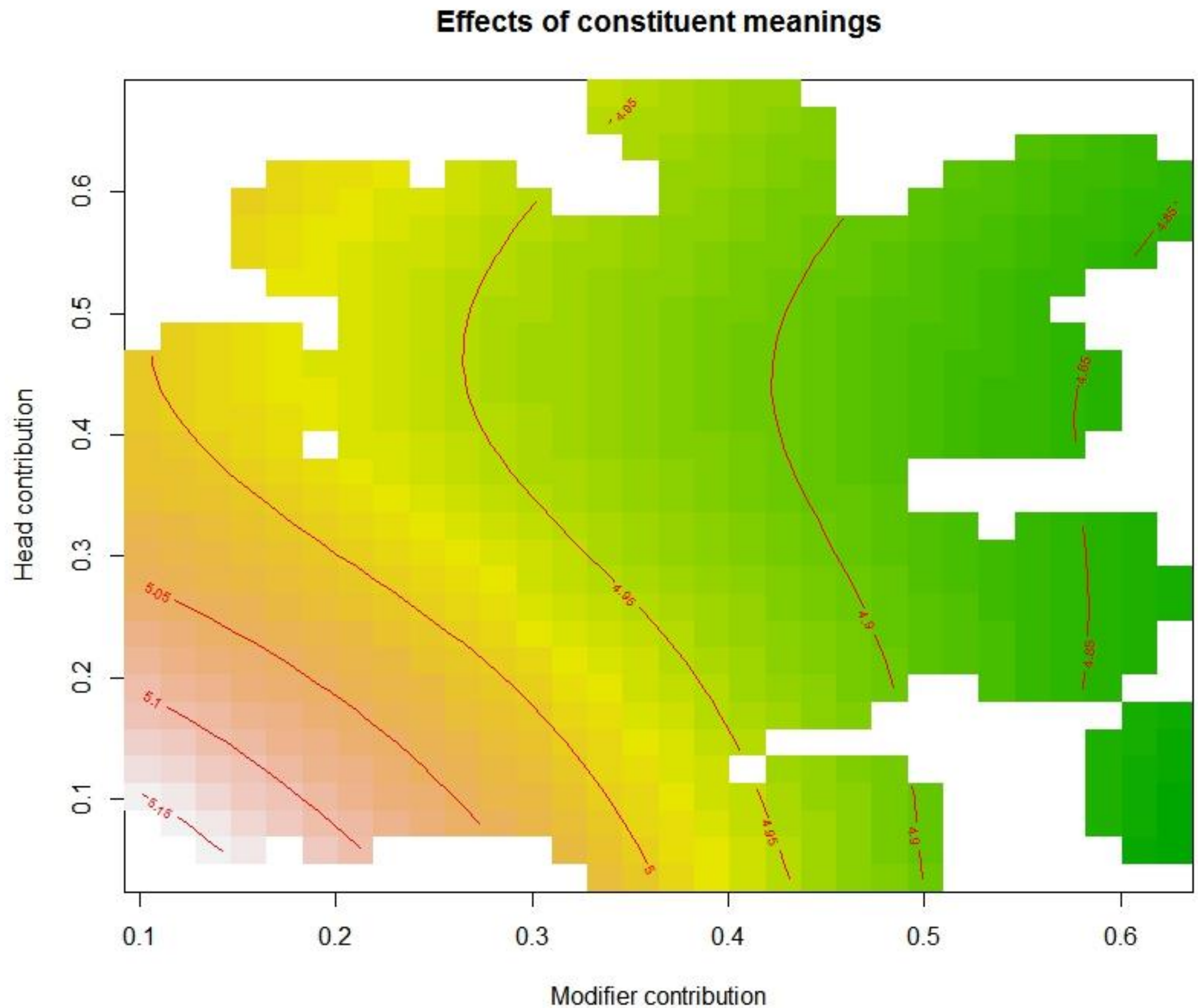


Fixation times on each word (ms)

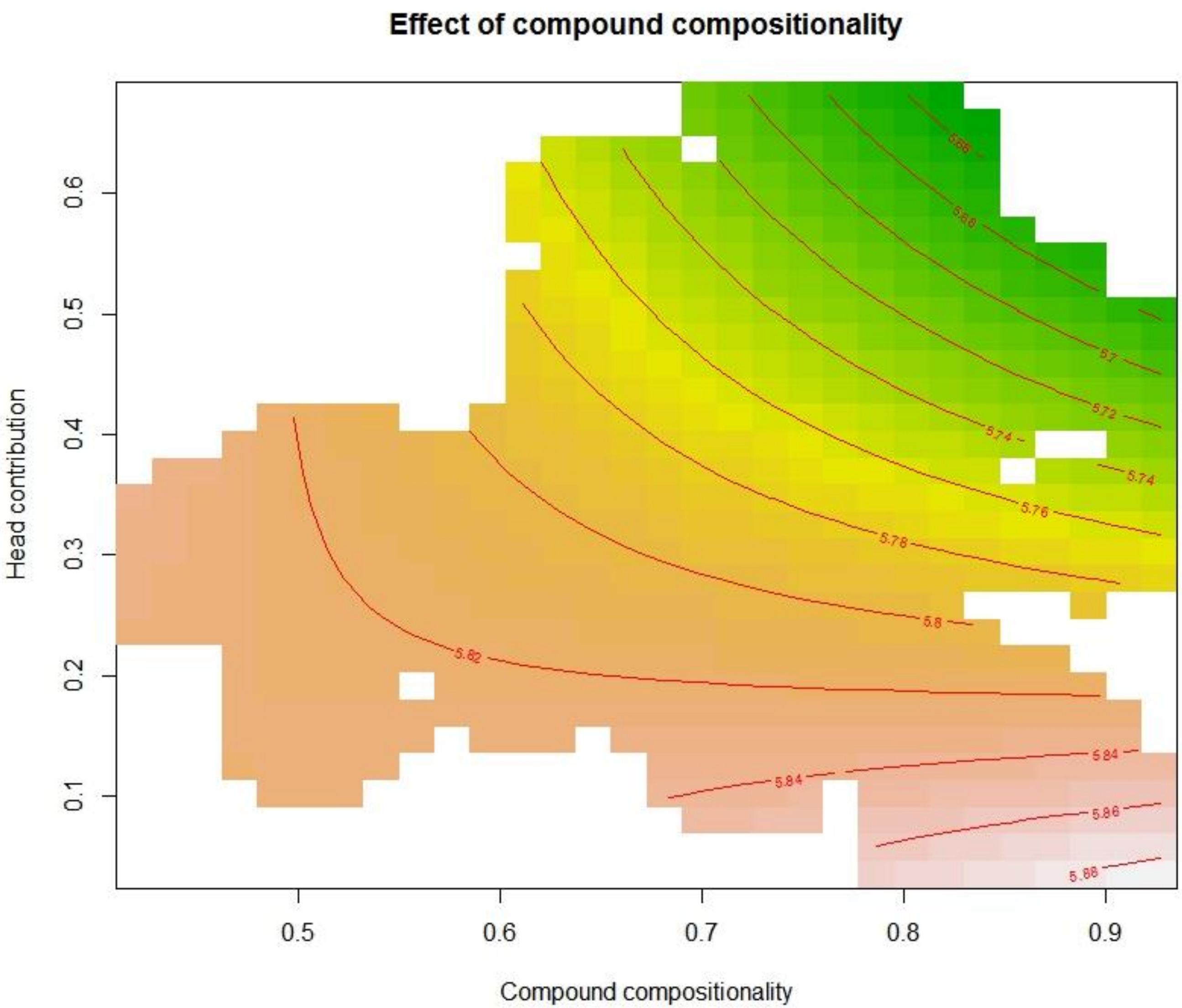
CAOSS effects in eye tracking



CAOSS effects on first fixations

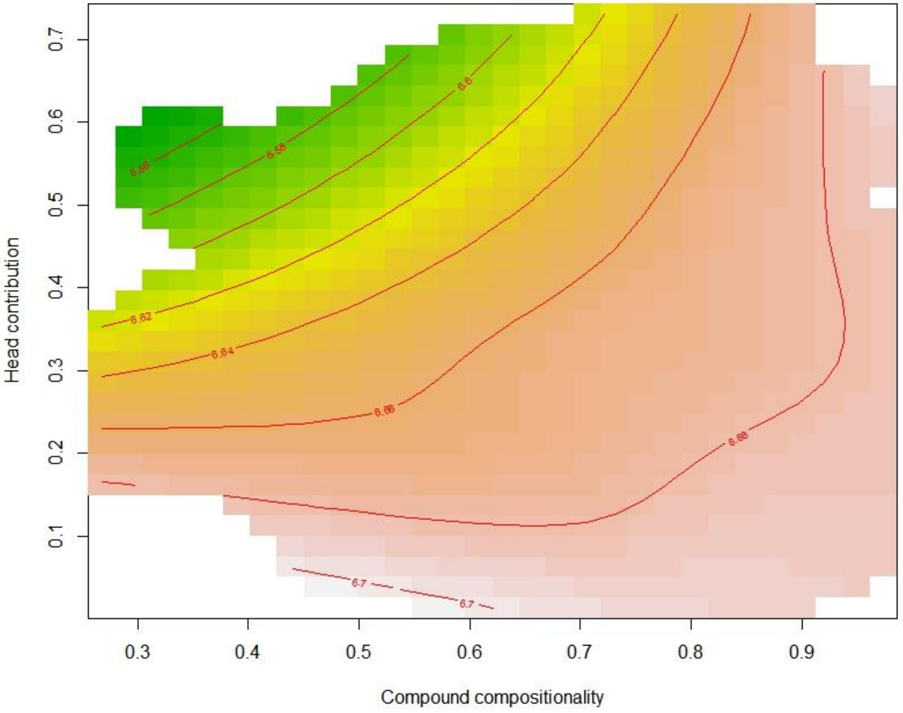


CAOSS effects on gaze durations

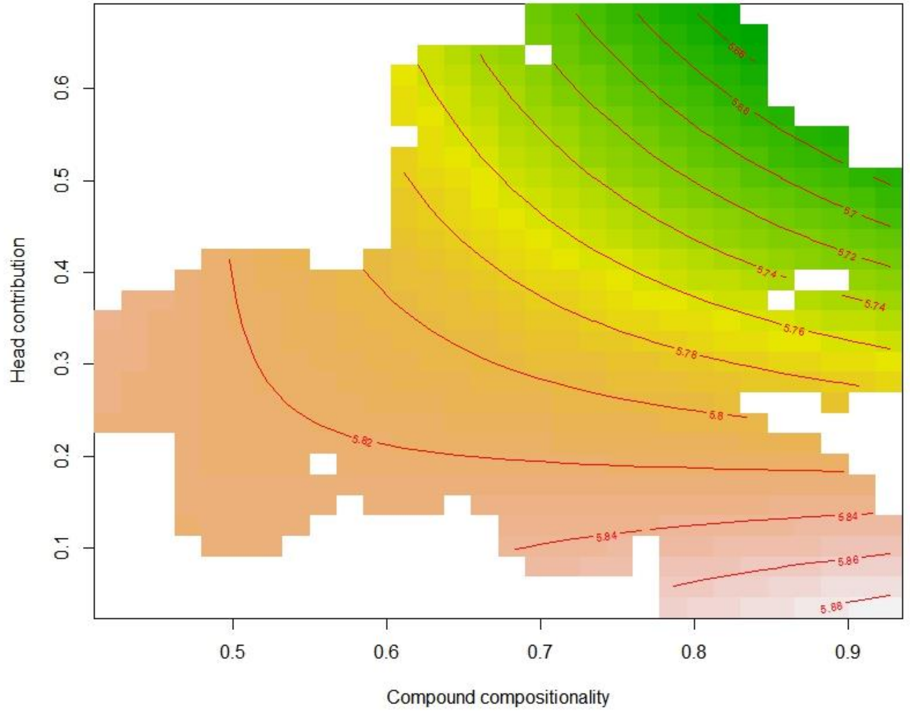


Compositionality and task effects

Lexical decision



Eye tracking in reading



CAOSS effects in eye tracking

- Time course of the compositional process
  - First, early combination of constituent meanings
  - Second, late comparison between compositional and stored compound meaning
- The effect of compound compositionality is affected by task requirements
  - When a specific sense must be accessed (reading task), a competition between the compositional and the lexicalized meaning needs to be resolved: compositionality eases the process

Conclusions

- There are complex semantic dynamics that must be formalized in order to be properly investigated
  - Distributional models can be profitably applied as a large-scale data-driven solution

- Compositionality plays a central role in compound processing
  - Novel and familiar compounds builds on the same basic processes
  - Compositionality must be properly addressed in psycholinguistic investigations on compounding

Thank you for your attention!

...and thanks to...

Marco Baroni

Christina Gagné and Thomas Spalding

Fritz Günther

...for their invaluable contribution
to the presented works