# Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space

by Baroni & Zamparelli (2010)

Johannes Wolf

Universität des Saarlands

Selected Topics in Semantics and Discourse

# Overview

# Syntax of Adjectives

- *The chair is* **green**.
  $\Rightarrow$ predicative
- *Peter is sitting on the* **green** *chair*
  $\Rightarrow$ attributive

# Syntax of Adjectives

Attributive adjectives

*chair*

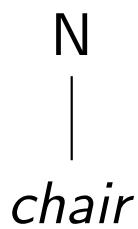
# Syntax of Adjectives

Attributive adjectives

*chair*

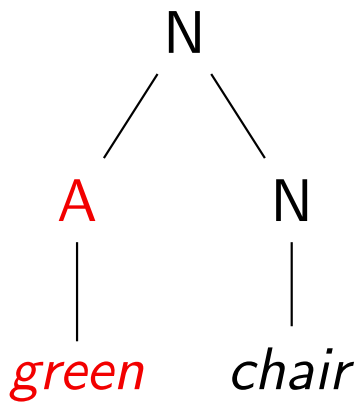
# Syntax of Adjectives

Attributive adjectives



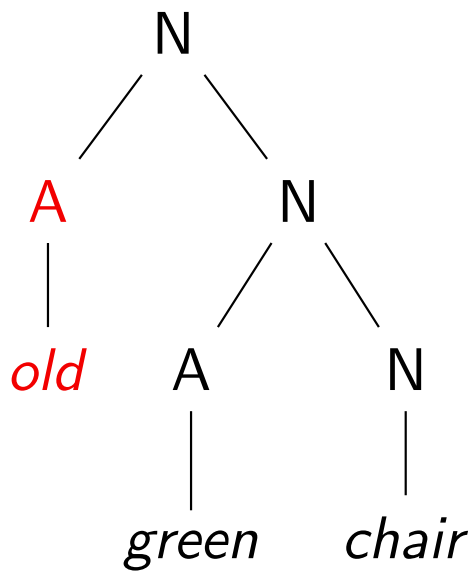
# Syntax of Adjectives

Attributive adjectives



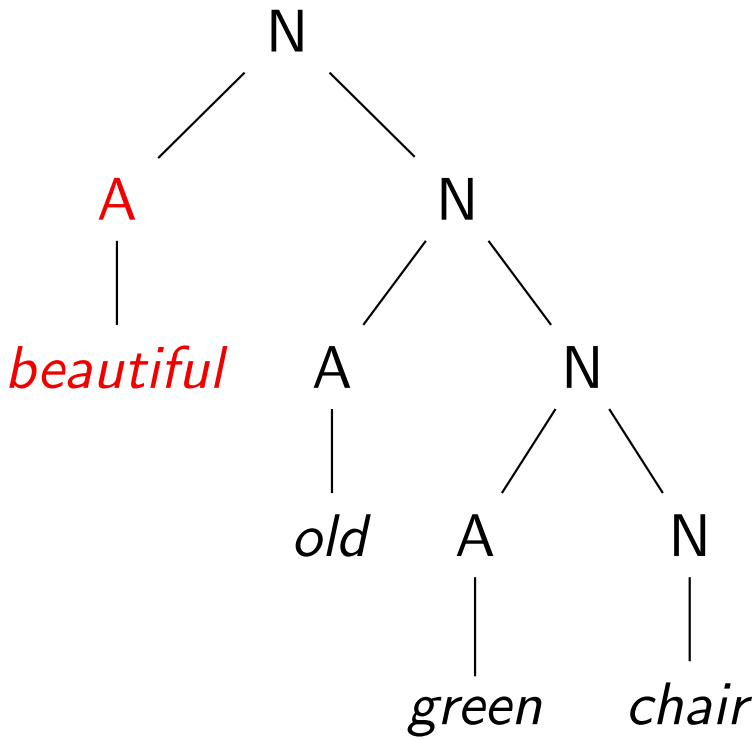
# Syntax of Adjectives

Attributive adjectives



# Syntax of Adjectives

Attributive adjectives



# Semantics of Adjectives: Distributive Semantics

- Semantic representation of a noun: $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$
- Semantic representation of an adjective: $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$
- Semantic representation of an AN (adjective + noun): $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

What is the semantic representation of *beautiful old green chair?*

# Semantics of Adjectives: Compositional Semantics (Montague)

- Semantic representation of a noun: **a set of individuals**
  meaning of student: set of all students
- Semantic representation of an adjective:
  a function from **noun-meanings** to **noun-meanings**

  How does the function work?

# Semantics of Adjectives: Compositional Semantics

**intersective** adjectives:
*blue, green, red, German, English, French, electric, plastic, ...*

- What is the meaning of *German*?
  a function that takes a set of individuals and returns a subset.
- meaning of *German student*:
  $\{\ldots \text{German individuals} \ldots\} \cap \{\ldots \text{students} \ldots\}$
- more complicated adjectives:
  *small, big, easy, ...*
  *fake, alledged, ...*

Note that compositional semantics has no difficulty predicting the meaning of *beautiful old green chair*.

# Compositional Distributive Semantics

Baroni & Zamparelli (2010):


- Meaning of noun: $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$
- Meaning of adjective: function from **(noun-)vectors** to **(noun-)vectors**
  How does this function work?

# Adjective-Noun Composition **(AN)**

Baroni & Zamparelli (2010): adjective-specific linear map

- the function for an adjective consists of an $n \times n$ matrix

$$\begin{pmatrix} y_{1,1} & \cdots & y_{1,n} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}^T$$

$$A \quad \cdot \quad N \quad = \quad AN_{approx}$$

# Experiment Design

- Corpus: ukWaC (.uk websites), Wikipedia, British National Corpus : 2.83B tokens
- AN test set: 36 adjectives across various classes
- nouns: 1,420 that occurred at least 300 times in post-adjectival position
- 26,440 ANs

# Experiment Design: Semantic Space

- rows: **core vocabulary** contains the top 8K most frequent noun lemmas and top 4K adjective lemmas
- columns: The 10K lemmas (nouns, adjectives or verbs) that co-occur with the largest number of items in the core vocabulary
- $$\begin{pmatrix} x_{1,1} & \cdots & x_{1,10,000} \\ \vdots & \ddots & \vdots \\ x_{12,000,1} & \cdots & x_{12,000,10,000} \end{pmatrix}$$

# Magic

MAGIC!
(Singular Value Decomposition)

# Experiment design: Semantic Space

- rows: **core vocabulary** contains the top 8K most frequent noun lemmas and top 4K adjective lemmas
- columns: The 10K lemmas that co-occur with the largest number of items in the core vocabulary
- $$\begin{pmatrix} x_{1,1} & \cdots & x_{1,300} \\ \vdots & \ddots & \vdots \\ x_{12,000,1} & \cdots & x_{12,000,300} \end{pmatrix}$$

# Evaluation: Study 1

Nearest 3 neighbors of centroids of ANs that share the same adjective

| *American N* | *black N* | *easy N* |
|---|---|---|
| Am. representative | black face | easy start |
| Am. territory | black hand | quick |
| Am. source | black (n) | little cost |
| *green N* | *historical N* | *mental N* |
| green (n) | historical | mental activity |
| red road | hist. event | mental experience |
| green colour | hist. content | mental energy |
| *necessary N* | *nice N* | *young N* |
| necessary | nice | youthful |
| necessary degree | good bit | young doctor |
| sufficient | nice break | young staff |

# Evaluation: Study 1

Nearest 3 neighbors of specific ANs

| *bad luck* | *electronic communication* | *historical map* |
|---|---|---|
| bad | elec. storage | topographical |
| bad weekend | elec. transmission | atlas |
| good spirit | purpose | hist. material |
| *important route* | *nice girl* | *little war* |
| important transport | good girl | great war |
| important road | big girl | major war |
| major road | guy | small war |
| *red cover* | *special collection* | *young husband* |
| black cover | general collection | small son |
| hardback | small collection | small daughter |
| red label | archives | mistress |

# Evaluation: Study 2

Comparison of different compositions models:

- last week: additive (add), multiplicative (mult) model
- this paper: adjective-specific linear map (alm)
- alternative: just stick with the meaning of noun or adjective
- model from Guevara (2010) (we will not talk about this)

# Evaluation: Study 2

Comparison of different compositions models:

- We take ANs that were not used fot training
- We list the closest neighbors of the predicted AN-vectors (only 1K neighbors per AN for efficiency)

| *method* | *25%* | *median* | *75%* |
|---|---|---|---|
| *alm* | 17 | 170 | $\geq$1K |
| *add* | 27 | 257 | $\geq$1K |
| *noun* | 72 | 448 | $\geq$1K |
| *mult* | 279 | $\geq$1K | $\geq$1K |
| *slm* | 629 | $\geq$1K | $\geq$1K |
| *adj* | $\geq$1K | $\geq$1K | $\geq$1K |

# Conclusions

Baroni & Zamparelli attempt to combine compositional and distributional semantics:

> *Ultimately, we want to compose larger and larger constituents, up to full sentences. It remains to be seen if the approach we proposed will scale up to such challenges.*

# Discussion

Thank you for your attention!

Questions?