From neighborhood to parenthood: the advantages of dependency representation over bigrams in Brown clustering

Simon Šuster and Gertjan van Noord
University of Groningen

Word representations

- Falls into following categories:
  - Word-space models (DS) + dimensionality reduction
  - Clustering
  - Word embeddings
  - Other probabilistic models
- Improve generalization
- Clustering: grouping similar words (semantic, paradigmatic & orthographic variants)

Brown clusters

- Popular: POS-tagging, NER, parsing, question answering etc.
- Easy-to-understand parameters
- Simplicity and robustness

- Word embeddings' recent momentum:
    - Brown clusters hard to beat in real tasks (Turian et al 2010; Bansal et al 2014; Nepal and Yates 2014; Passos et al 2014)
- Brown clusters admittedly less scalable

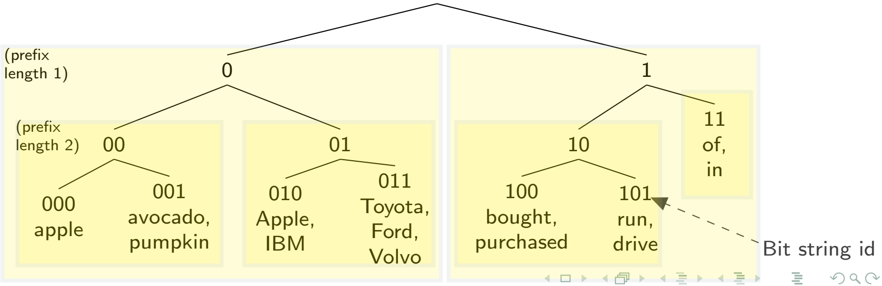
Our contribution

- Original Brown clustering uses bigram contexts
- Adapt to dependencies: helpful for semantic similarity
- Tool for dependency Brown clustering

How does Brown clustering work

- Maximize data likelihood defined on a class-based bigram LM
- In practice, done through average Mutual Information

1. assign some word types to unique classes
2. put each remaining word type to one of these classes by minimizing the MI loss
3. when all word types are merged, further merge the resulting classes to create a hierarchy



From Brown to dependency Brown

Original formulation

- Factorization includes class transitions with conditioning on *previous* word's class
- Such representation is *local*

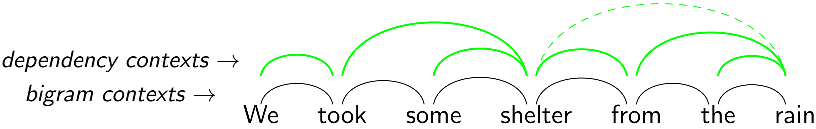
Our modification

- Adopt dependency representation: less local, more precise (assuming we can trust the parser)
- (Class-based) dependency language model: conditioning on *parent's* class

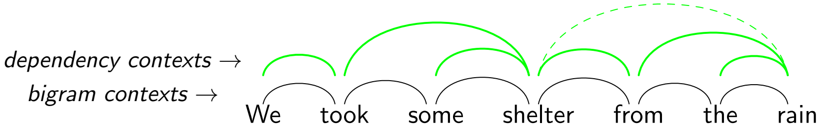
Contexts

bigram contexts → We took some shelter from the rain

Contexts


dependency contexts →
bigram contexts →
We took some shelter from the rain

Contexts



Extract parent–child pairs:
(*took*, *We*),
(*took*, *shelter*),
(*shelter*, *some*),
...
(Optional) 2nd order dependency: collapse on preposition
(shelter, rain)

Evaluation

- Parse a 46M-word reference corpus sample
- Obtain counts of dependency instances as input for the clustering tool
- Semantic similarity task on Dutch wordnet
  - Average similarity over all clusters, as measured by Lin score

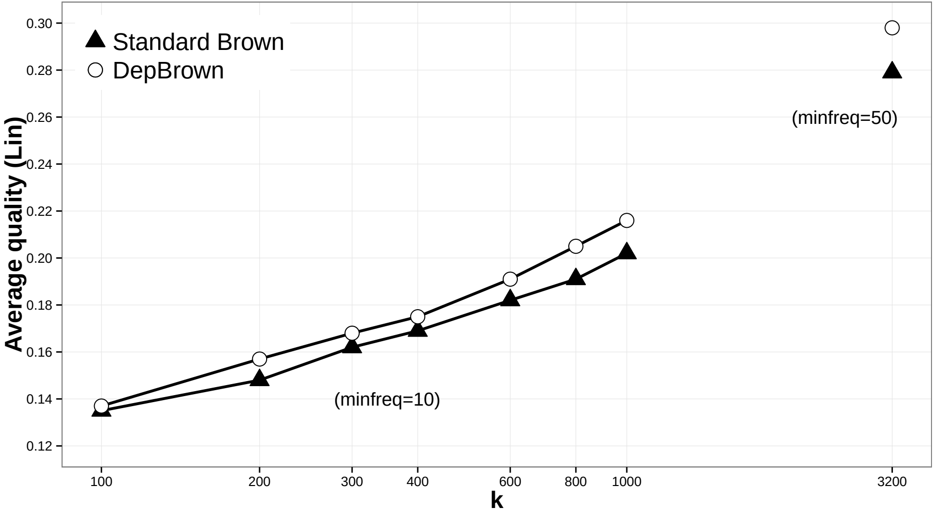
Group	Cluster id	Most frequent words
A1	001010001011100	contractor, family doctor, baker, lawyer, pharmacist, real estate agent, property developer, postman, ...
A2	001010001011011	analyst, reviewer, observer, expert, commentator, people's rights organisation, insider, ...
A3	0010100010111110	entrepreneur, businessman, manager, self-employed, merchant, starter, craftsman, ...
B1	011101111011110	me
B2	01110111101110	him/herself, myself, yourself
B3	01110111101100	them
C1	00110010010	Bush, Obama, Clinton, Putin, ...
C2	0011000111010	Sarah, Kim, Nathalie, Justine, ...
C3	0011000111011	David, Jimmy, Benjamin, ...
D1	001011100010101	email, mail, sms, sms_DIM, e-mail, mail_DIM, ...
D2	001011100010100	telephone, satellite, telephony, telephone line, Explorer, music player, iTunes, ...
E	001000010110101	income, energy consumption, minimum wage, cholesterol, IQ, alcohol content, ...

(translated from Dutch)

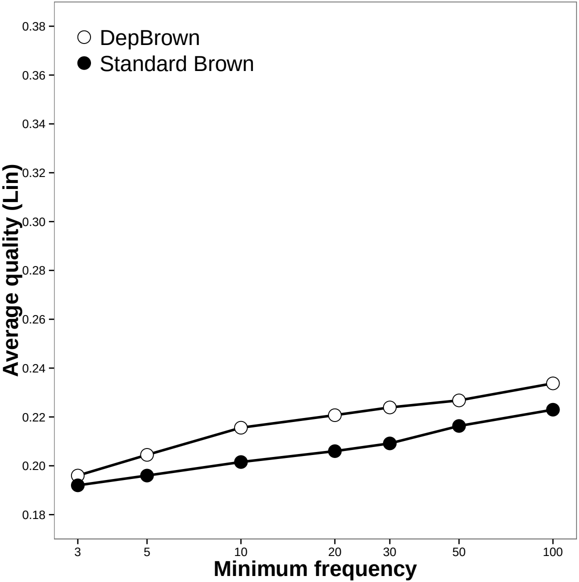
Different contexts, different clusters?

- No strong evidence in our case (manual inspection)
- Word embedding and distributional-semantic literature:
  - Bow: words associated with target word (topical similarity)
  - Dep: words behaving like target word
- Bigram contexts in original Brown clustering too narrow for topical similarity

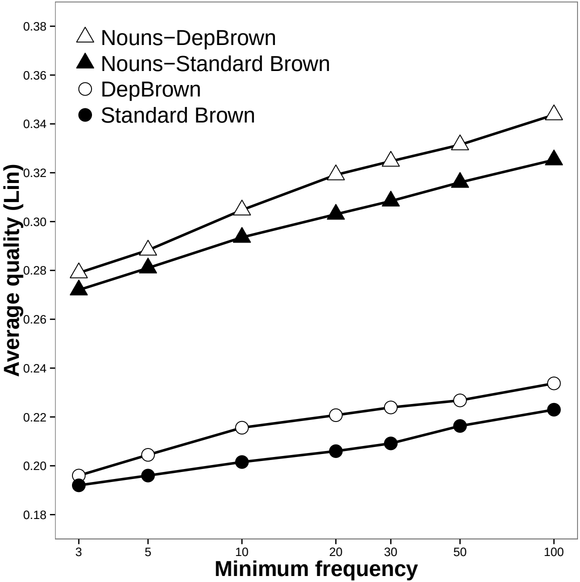
Varying *k* number of clusters



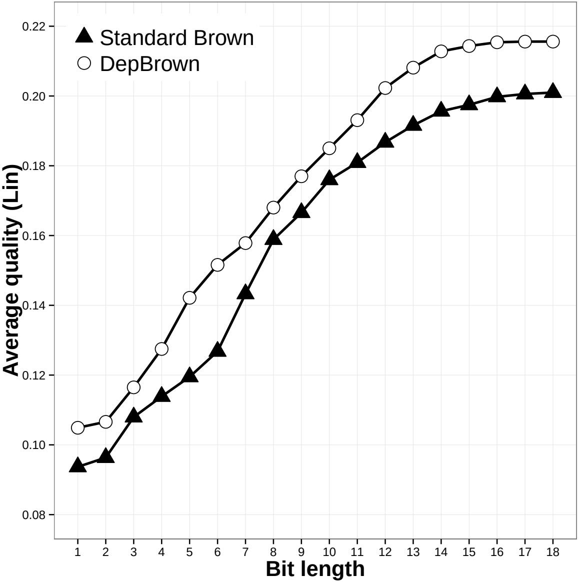
Varying *minfreq*



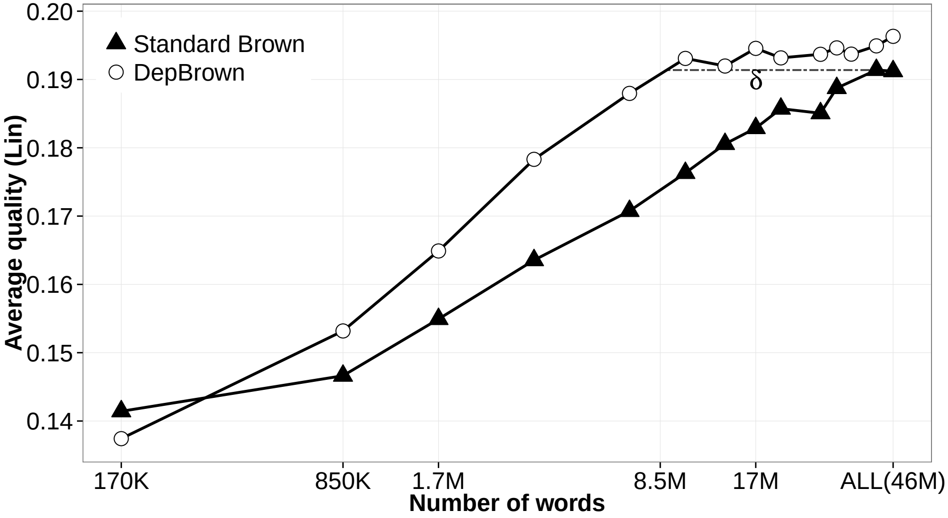
Varying *minfreq* + Nouns only



Prefix length



Amount of data



Leveraging syntactic functions

- Select parent–child pairs based on dependency label
- Further improvements in semantic similarity by using:
  - subjects
  - direct objects
  - directional complements
  - 2-nd order relations (intervening preposition)
    - directional and prepositional complements

Thank you!

Data & clustering tool at:

github.com/rug-compling/dep-brown-cluster