Computational Semantics and Pragmatics

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam

Autumn 2016

Linguistic coordination

Three different approaches to explaining coordination of linguistic forms:

- driven by *communicative* goals and the need for mutual understanding
- consequence of our *cognitive* architecture, triggered by priming mechanisms
- driven by *social* goals, to negotiate social distance

To read for discussion today :

C. Danescu-Niculescu-Mizil, L. Lee, B. Pang and J. Kleinberg (2012). Echoes of power: Language effects and power differences in social interaction, *Proceedings of WWW*.

Reitter & Moore (2014), Alignment and task success in spoken dialogue, *Journal of Memory and Language*.
↪ Shorter older version: Reitter & Moore (2007). Predicting Success in Dialogue, ACL.

C. Danescu-Niculescu-Mizil, L. Lee, B. Pang and J. Kleinberg (2012). Echoes of power: Language effects and power differences in social interaction, *Proceedings of WWW*.

1. Which forms of power are considered?

2. Which domains/datasets are are used to investigate these forms of power?

3. What are the theoretical assumptions / hypotheses tested?

4. What's the measure of linguistic style coordination used?

5. What are the results for each domain and form of power?

Style Coordination

C. Danescu-Niculescu-Mizil, L. Lee, B. Pang and J. Kleinberg (2012). Echoes of power: Language effects and power differences in social interaction, *Proceedings of WWW*.

How things are said as opposed to *what* is said

⇝ *function words* are topic-independent (Pennebaker et al, 2007)
pronouns, articles, quantifiers, prepositions, conjunctions, . . .

> Editor$_a$: Corrected. Please check. **Any** more outstanding problems?
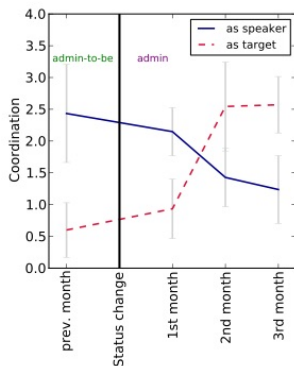> Editor$_b$: **Everything** is fine. Thanks a lot.

Coordination of b towards a for a class of function words m, for all pairs of utterances (u_a, u_b) where b directly replies to a:

$$C^m(b, a) = P(u_b \text{ uses } m \mid u_a \text{ used } m) - P(u_b \text{ uses } m)$$

Status-based power : Wikipedia editors coordinate more towards admins (Danescu-Niculescu-Mizil et al., 2012. *Echoes of Power.*) ▸ website

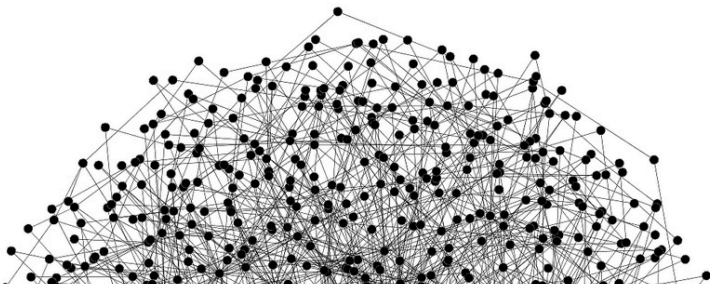Status change has an effect on the level of linguistic style coordination:

Coordination of the user (as speaker) and, respectively, towards the user (as target) in the months before and after status change occurs.

Power-Driven Style Coordination

Status-based power : Wikipedia editors coordinate more towards admins (Danescu-Niculescu-Mizil et al., 2012. *Echoes of Power.*)

What about other more implicit forms of social power, such as how *'central'* you are within the *social network* – do they impact linguistic style matching?

B. Noble and R. Fernández (2015). Centre Stage: How Social Network Position Shapes Linguistic Coordination. *Workshop on Cognitive Modeling and Computational Linguistics, NAACL 2015.*
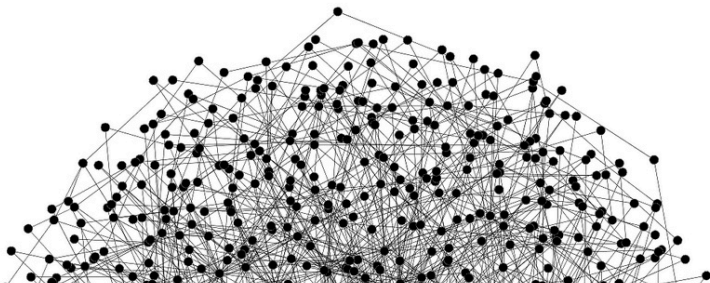
The Wikipedia Social Network

We want to construct a *social network* that reflects the *linguistic interactions* between the Wikipedia editors:
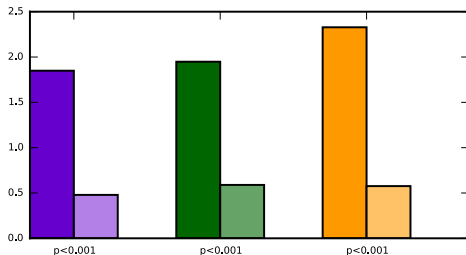
- nodes represent individuals in a community – Wikipedia editors
- edges give some measure of social connectivity between individuals
  – weighted according to the number of direct replies

Corpus: 342,800 posts, 26,397 editors (1,825 of whom are admins)

Centrality Measures

Betweenness centrality: How important are you to community connectivity?

$$\mathrm{BC}(n^*) = \sum_{n \neq m \in N} \frac{|\{\sigma \in \mathrm{Path}(m,n) \mid n^* \in \sigma\}|}{|\mathrm{Path}(m,n)|}$$

where $\mathrm{Path}(m,n)$ is the set of shortest paths between m and n

Eigenvector centrality: How important are your neighbours?

$$\mathrm{EC}(n^*) = \frac{1}{\lambda} \sum_{n \in M(n^*)} \mathrm{EC}(n)$$

where $M(n)$ is the neighbourhood of n and λ is the largest *eigenvalue*

Highly central editors: over one standard deviation above mean score.

Results

- More style coordination towards administrators.
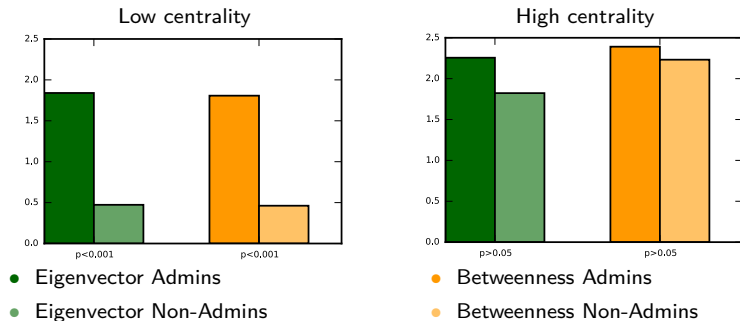- More style coordination towards editors in central social positions.

- Admins
- Non-Admins
- High Eigenvector
- Low Eigenvector
- High Betweenness
- Low Betweenness

- On average, admins occupy more central positions, but the impact of adminship and centrality turn out to be largely independent ...

Results

- Low-centrality editors receive more coordination if they are admins.
- But adminship is less important for high-centrality users.

Low centrality

- Eigenvector Admins
- Eigenvector Non-Admins

High centrality

- Betweenness Admins
- Betweenness Non-Admins

⤳ *social network centrality* sometimes eclipses status-based power in triggering linguistic style adaptation.

Open Issues . . .

Is adaptation to central users (rather than admins) more important for *social acceptance*?

- how does this happen even though centrality is more *implicit*?
- do highly central users exhibit speech *more typical* of the community?

Several *practical applications* within computational social science:

- automatic discovery of social relations,
- tracking evolution of relations over time, . . .

1. What does the paper investigate? what is the theoretical motivation? what hypotheses are considered?

2. Which data is used? how is task success determined?

3. What kind of linguistic alignment is investigated?

4. Which experiments are performed? how is alignment/priming measured in these experiments?

5. What are the results? what are their implications according to the authors?

Research projects

Submit a project proposal by Thursday 13 October at 14:00.
[Between 250 and 500 words, PDF or plain text, sent to raquel by email]

- *Who:* team members (ideally two).
- *What:* which phenomenon do you want to investigate, what is your research question?
- *Why:* provide some motivation, why is this interesting? do you have specific hypotheses?
- *How:* very important part!
  - which data and methods will you use?
  - which steps do you foresee? is this feasible given the time constraints? make a plan.

We will discuss the proposals on Fri 14 Oct.

Research projects

Some dialogue corpora freely available or available at ILLC:

- Switchboard
- British National Corpus (BNC)
- Corpus gesproken Nederlands
- Wikipedia talkpage corpus
- Film corpus
- Internet Argument Corpus
- PentoRef (push-to-talk vs free turn taking)
- CHILDES