Polytomous logistic regression analysis and modeling of linguistic alternations

Antti Arppe
General Linguistics, Department of Modern Languages
University of Helsinki
Concepts – linguistic alternations

- Alternative linguistic **forms** which denote roughly the same **meaning**
  - **Structural/constructional** alternations
    - E.g. Finnish/German word order, English dative (Bresnan 2007) or possessive alternations (Gries 2003)
      - *He gave her the book* vs. *He gave the book to her*
      - *The book’s title* vs. *the title of the book*
  - **Lexical** alternations
    - E.g. (near-)synonymy, social/dialectal variation
      - *Strong* vs. *powerful* (Church et al. 1991)
      - *Small* vs. *wee*
Theoretical assumptions & methodological prerequisites

- *Monocausal/univariate* explanations of linguistic phenomena are insufficient or contradictory (e.g. Gries 2003a)
  - Lexical or syntactic choices made by speakers are determined, and can thus be explained by a *plurality* of factors, in *interaction*
  - Necessity of *multifactorial* explanatory models → *multivariate* statistical analysis
Theoretical assumptions & methodological prerequisites

- **Probabilistic grammar**
  - Bod et al. (2003) and Bresnan (2007) have suggested that the selections of alternative selections on context, i.e. outcomes for combinations of variables, are generally speaking **probabilistic**
  - even though the individual choices in isolation are **discrete**

- In other words, the workings of a linguistic **system**, represented by the range of variables according to some **theory**, and its resultant **usage** are
  - in practice **not categorical**, following from **exception-less** rules,
  - but rather exhibit degrees of **potential variation** which becomes evident over **longer stretches of linguistic usage**
  - Integral characteristic of language – **not** a result of “interference” from language-external cognitive processes
Discrete vs. probabilistic

- ... 
- \(XAY\) 
- \(YBX\) 
- \(XAY\) 
- \(XAY\) 
- \(XAY\) 
- \(XAY\) 
- \(YBX\) 
- \(XCY\) 
- ... 

- \(X\_Y\)
  - A: 4
  - C: 1
- \(Y\_X\)
  - B: 2
- \(X, Y\)
  - A: 5
  - B: 2
  - C: 1
Discrete vs. probabilistic – Interpretation of the previous data

- If we assume categorical rules, can we extract them?
  - $Y_X \rightarrow B$
  - $X_Y \rightarrow A?/C$?
  - $X,Y \rightarrow A?/B?/C$

- What do we assume about the nature of these rules and their relationship with the data?
  - Is e.g. feature order a permissible or truly relevant characteristic?
    - $Y_X \rightarrow B \sim X_Y \rightarrow B$?
  - Do we expect that some additional variables (e.g. extralinguistic or stylistic) – yet unnoticed – might explain away the remaining irregularities?
    - $X_Y \rightarrow A$
    - $X_YW \rightarrow C$
  - Can we explain all cases exhaustively and categorically by adding new explanatory variables?
- Or do we rather allow *a priori* for variation and proportionate occurrence in the scrutinized contexts
  - X,Y ->
    - A (62.5%) |
    - B (25%) |
    - C (12.5%)

Figure 1: Sample probabilities from the corpus model of Bresnan et al.
Theoretical assumptions & methodological prerequisites

- **Polytomous** vs. **dichotomous** linguistic alternations: often more than two alternatives (cf. Divjak & Gries 2007; any [synonym] dictionary)
  - Structural alternation: English relative clauses
    - The book *which* I read was good.
    - The book *that* I read was good.
    - The book [] I read was good.
  - Lexical alternations: (English) synonyms
    - Do you *understand* what I mean?
    - Do you *comprehend* what I mean?
    - Do you *grasp* what I mean?
    - Do you *get* what I mean?
    - ....
Lexical alternation – practical example case

- Set of the most frequent synonyms denoting THINK in Finnish
  - \textit{ajatella} < \textit{ajaa} 'to drive habitually (in one’s mind)'
  - \textit{miettiä} < \textit{sметит’} Slavic (Baltic?) loan to the Fennic languages (i.e. 2000-3000 years old) cf. Swedish/Germanic \textit{mäta} 'to measure'
  - \textit{pohtia} ~ \textit{pohtaa} < archaic/agricultural (→1950s) 'to winnow'
  - \textit{harkita} < \textit{harkki} archaic/agricultural 'dragnet' ~ \textit{haroa/haravoida} 'to rake'
  - \textit{[tuumia/tuumata} < Russian \textit{dumat’} 'to think' (Slavic loan) cf. Swedish/Scandinavian \textit{dömma} 'to judge, deem']

- Currently translatable into English as:
  - 'think, reflect, ponder, consider'
Research corpus – two sources

- Two months worth (January–February 1995) of written text from Helsingin Sanomat (1995)
  - Finland's major daily newspaper
  - 3,304,512 words of body text
  - excluding headers and captions, as well as punctuation tokens
  - 1,750 representatives of the studied THINK verbs

- Six months worth (October 2002 – April 2003) of written discussion in the SFNET (2002-2003) Internet discussion forum, namely regarding
  - (personal) relationships (sfnet.keskustelu.ihmissuhteet)
  - politics (sfnet.keskustelu.politiikka)
  - 1,174,693 words of body text
  - excluding quotes of previous postings as well as punctuation tokens
  - 1,654 representatives of the studied THINK verbs

- The proportion of the THINK lexemes in the Internet newsgroup discussion text is more than twice as high as the corresponding value in the newspaper corpus

- The individual overall frequencies among the studied THINK lexemes in the research corpora were
  - 1492 for ajatella
  - 812 for miettiä
  - 713 for pohtia
  - 387 for harkita
Selected on the basis of extensive univariate analysis

- Altogether 48 contextual feature variables:
  - Morphological features pertaining to the node-verb or the entire verb-chain they are components of (10)
  - semantic characterizations of verb-chains (6)
  - syntactic argument types, without any subtypes (10)
  - Syntactic arguments combined with their semantic and structural subtypes (20)
  - extra-linguistic features (2)
Overall model

- \{ajatella|miettiä|pohtia|harkita}\ ~ Z\_ANL\_NEG + Z\_ANL\_IND + Z\_ANL\_KOND + Z\_ANL\_PASS + Z\_ANL\_FIRST + Z\_ANL\_SECOND + Z\_ANL\_THIRD + Z\_ANL\_PLUR + Z\_ANL\_COVERT + Z\_PHR\_CLAUSE + SX\_AGE\_SEM\_IN\_GROUP + SX\_PAT\_SEM\_GROUP\_INDIVIDUAL + SX\_PAT\_SEM\_GROUP\_ABSTRACTION + SX\_PAT\_SEM\_GROUP\_ACTIVITY + SX\_PAT\_SEM\_EVENT + SX\_PAT\_SEM\_COMMUNICATION + SX\_PAT\_INDIRECT\_QUESTION + SX\_PAT\_DIRECT\_QUOTE + SX\_PAT\_ + SX\_LX\_että\_CS\_SX\_PAT + SX\_SOU + SX\_GOA + SX\_MAN\_SEM\_GENERIC + SX\_MAN\_SEM\_FRAME + SX\_MAN\_SEM\_POSITIVE + SX\_MAN\_SEM\_NEGATIVE + SX\_MAN\_SEM\_AGREEMENT + SX\_MAN\_SEM\_JOINT + SX\_QUA + SX\_LOC + SX\_TMP\_SEM\_DEFINITE + SX\_TMP\_SEM\_INDEFINITE + SX\_DUR + SX\_FRQ + SX\_META + SX\_RSN\_PUR + SX\_CND + SX\_CV + SX\_VCH\_SEM\_POSSIBILITY + SX\_VCH\_SEM\_NECESSITY + SX\_VCH\_SEM\_EXTERNAL + SX\_VCH\_SEM\_VOLITION + SX\_VCH\_SEM\_TEMPORAL + SX\_VCH\_SEM\_ACCIDENTAL + Z\_EXTRA\_SRC\_sfnet + Z\_QUOTE
Selection of multivariate statistical method

- **Logistic regression** – *WHY?*
  - Looks at outcomes as proportions among all observations with the same context
    - rather than individual *either-or* dichotomies of occurrence vs. non-occurrence
    - Thus estimates *probabilities of occurrence* given a particular context
    - Thus, also compatible with the probabilistic view of language
  - Estimates variable parameters which can be interpreted “naturally” as *odds* (Harrell 2001)
    - How much does the existence of a variable (i.e. feature) in the context increase (or decrease) the *chances* of a particular outcome (i.e. lexeme) to occur, with all the other explanatory variables being equal?
Logistic regression – formalization of binary (dichotomous) setting

- Model X with M explanatory variables \( \{X\} \) and parameters \( \{\alpha_k, \beta_k\} \) for outcome \( Y=k \):
  \[
  X = \{X_1, \ldots, X_M\}
  \]
  \[
  \beta_k X = \beta_{k,1}X_1 + \beta_{k,2}X_2 + \ldots + \beta_{k,M}X_M
  \]
  \[
  P_k(X) = P(Y=k|X); \ P_{\neg k}(X) = P(Y=\neg k|X) = 1-P(Y=k|X)
  \]

- \( \text{logit}[P_k(X)] = \text{log}_e\{P_k(X)/[1-P_k(X)]\} = \alpha_k + \beta_k X \)
  \[
  \iff P_k(X)/[1-P_k(X)] = \exp(\alpha_k + \beta_k X)
  \]
  \[
  \iff P_k(X)/[1-P_k(X)] = \exp(\alpha_k) \cdot \exp(\beta_k X)
  \]
  \[
  = \exp(\alpha_k) \cdot \exp(\beta_{k,1}X_1) \cdot \ldots \cdot \exp(\beta_{k,M}X_M)
  \]
  \[
  \iff P_k(X) = 1/[1+\exp(-\alpha_k - \beta_k X)]
  \]
Binary logistic regression – a concrete example …

\[ Miten_{MANNER+GENERIC} \ajatteli_{INDICATIVE+SECOND, \ COVERT, \ AGENT+INDIVIDUAL} \erota_{PATIENT+INFINITIVE} \ldots \jostain \ldots \text{SAKn kannattajasta?} \text{[sfnet]} \]

‘How did you think to differ at all from some dense supporter of class-thinking in SAK?’

Context \( \subset \mathcal{X} = \{MANNER:GENERIC, \ INDICATIVE, \ SECOND\_PERSON, \ COVERT\_AGENT, \ AGENT:INDIVIDUAL, \ PATIENT:INFINITIVE, \ SFNET\} \)
Binary logistic regression – a concrete example …

*Miten* MANNER+ GENERIC *ajattelit* INDICATIVE+ SECOND, COVERT, AGENT+ INDIVIDUAL *eroa* PATIENT+ INFINITIVE … *jostain … SAKn kannattajasta?* [sfnet]

‘How did you **think** to differ at all from some dense supporter of class-thinking in SAK?’

\[
\log_e \left[ \frac{\Pr(\text{ajatella} | \text{Context})}{\Pr(\neg \text{ajatella} | \text{Context})} \right] = 0.5 \approx \log_e \left[ \frac{3404-1492}{3404} \right] = 3:2 \]

\[\begin{array}{ll}
+3.0 \sim \text{MANNER: GENERIC} & \cdot (41:2) \sim \text{MANNER: GENERIC} \\
+0.6 \sim \text{INDICATIVE} & \cdot (13:7) \sim \text{INDICATIVE} \\
-(0.5) \sim \text{SECOND PERSON} & \cdot (1:2) \sim \text{SECOND PERSON} \\
+(0.0) \sim \text{COVERT SUBJECT} & \cdot (1:1) \sim \text{COVERT SUBJECT} \\
-(0.2) \sim \text{AGENT: INDIVIDUAL} & \cdot (5:6) \sim \text{AGENT: INDIVIDUAL} \\
+(1.8) \sim \text{PATIENT: INFINITIVE} & \cdot (6:1) \sim \text{PATIENT: INFINITIVE} \\
+(0.5) \sim \text{INTERNET-GENRE} & \cdot (3:2) \sim \text{INTERNET-GENRE} \\
\end{array}\]

\[\approx +5.8 \quad \approx +5.8 \quad \approx 319:1\]
Binary logistic regression – another concrete example ...

Vilkaise \text{CO-ORDINATED\_VERB}(+\text{MENTAL}) joskus \text{FREQUENCY}(+\text{SOMETIMES}) valtuuston esityslistaa ja mietsi \text{(IMPERATIVE+)} \text{SECOND,COVERT, AGENT+INDIVIDUAL monestako} \text{PATIENT+INDIRECT\_QUESTION} asiasta sinulla on jotain tietoa. [sfnet]

‘Glance sometimes at the agenda for the council and \textbf{think} on how many issues you have some information.’

\[
\begin{align*}
\log_e \left[ \frac{P(\text{miettiä}|\text{Context})}{P(\neg\text{miettiä}|\text{Context})} \right] &\approx -2.0 \approx \log_e \left( \frac{812}{3404} \right) \\
&= \frac{12.6}{1+12.6} \\
&\approx 0.93 \Rightarrow 0.88
\end{align*}
\]

\[
\begin{align*}
\log_e \left[ \frac{P(\text{miettiä}|\text{Context})}{P(\neg\text{miettiä}|\text{Context})} \right] &\approx 2.5 \\
&= \frac{12.6}{1+12.6} \\
&\approx 0.93 \Rightarrow 0.88
\end{align*}
\]
Binary logistic regression – still another concrete example …

Tarkastusviraston mielestä täätä ehdotusta olisi syytä pohtia tarkemmin.

‘In the opinion of the Revision Office there is reason to ponder this proposal more thoroughly.’

\[ P(\text{pohtia}|\text{Context})/P(\neg\text{pohtia}|\text{Context}) = 1:5 \sim \text{Intercept} (\approx 719/3404) \]
\[ \cdot (3:4) \sim \text{META-COMMENT} \]
\[ \cdot (4:3) \sim \text{PATIENT:ACTIVITY} \]
\[ \cdot (4:5) \sim \text{CONDITIONAL (MOOD)} \]
\[ \cdot (8:9) \sim \text{THIRD_PERSON} \]
\[ \cdot (8:9) \sim \text{COVERT_AGENT} \]
\[ \cdot (1:1) \sim \text{VERB-CHAIN:NECESSITY} \]
\[ \cdot (5:6) \sim \text{MANNER:SUFFICIENT} \]
\[ \approx 4:33 \approx 0.122:1 \approx 1:8.2 \]

\[ P(\text{pohtia}|\text{Context}) = 0.12/(1+0.12) \sim 0.11 (\rightarrow 0.125) \]
Binary logistic regression – still another concrete example …

In the opinion of the Revision Office there is reason to ponder this proposal more thoroughly.

\[
P(harkita|\text{Context})/P(\neg harkita|\text{Context}) = 4:41 \sim \text{Intercept} (\approx 387/3404) = 12/(1+12) \approx 0.92 (\Rightarrow 0.725)
\]

\[
\begin{align*}
&\cdot 3:2 \sim \text{META-COMMENT} \\
&\cdot 23:3 \sim \text{PATIENT:ACTIVITY} \\
&\cdot 14:5 \sim \text{CONDITIONAL (MOOD)} \\
&\cdot (22:15) \sim \text{THIRD_PERSON} \\
&\cdot (7:8) \sim \text{COVERT_AGENT} \\
&\cdot (10:7) \sim \text{VERB-CHAIN:NECESSITY} \\
&\cdot (2:1) \sim \text{MANNER:SUFFICIENT} \\
&\approx 12:1
\end{align*}
\]
Model fit – observed proportions vs. estimated probabilities

- Most frequent feature combination in data:
  - \( n\{Z\_ANL\_IND, Z\_ANL\_THIRD, SX\_AGE\_SEM\_INDIVIDUAL, SX\_PAT\_DIRECT\_QUOTE\}=88 \)

- Observed frequencies
  
<table>
<thead>
<tr>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>31</td>
<td>57</td>
<td>0</td>
</tr>
</tbody>
</table>

- Observed proportions
  
<table>
<thead>
<tr>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.35</td>
<td>0.65</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- Estimated probabilities
  
<table>
<thead>
<tr>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>0.37</td>
<td>0.60</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Dichotomous → Polytomous setting

- Example case: four outcomes (i.e. synonyms)
  - \{ajatella, miettiä, pohtia, harkita\}

- How could the selection of these be broken down into a set of binary models?
  - N.B. *nnet:multinom* consists of binary models!
Polytomous outcome setting – *binarization* techniques
Several heuristic techniques for binarizing (dichotomizing) polytomous outcome settings

- Baseline-category multinomial
  - simultaneously/separately fit
- One-vs-rest (one-against-all)
- Pairwise contrast (all-against-all, round-robin)
- Nested dichotomy
- Ensemble of nested dichotomies (ENDs)
Characteristic dimensions of polytomous logistic regression heuristics

- Number of constituent binary logistic regression models (→ complexity)
- Interpretation of explanatory variables in model(s) as well as the associated odds
  - Outcome-specific odds?
- Direct probability estimates for outcomes?
  - Necessity of normalization?
- Selection algorithm in prediction
Baseline-category multinomial

- **Reasoning**: one outcome is (manually/automatically) selected as a baseline category (most frequent, prototypical, or general), against which the other outcomes are contrasted each individually (Cox 1958)
  - Binary models may be fitted separately or dependently
- \{ajatella vs. miettiä\}, \{ajatella vs. pohtia\}, \{ajatella vs. harkita\}
- Variables and associated odds contrast other outcomes only with baseline (and not with each other)
- Number of binary models: n(outcomes)–1
- Direct probability estimates:
  - \( P(\text{baseline outcome}) = 1-\Sigma P(\text{non-baseline outcomes}) \)
  - Normalization of probabilities required, so that \( \Sigma P(\text{all outcomes})=1 \)
Baseline-category multinomial

(3.17) $P_k(X) = P(Y=k|X)$, with $\sum_{k=1}^{K} P_k(X) = 1$ and $k = \{1, \ldots, K\}$, and $P_K(X) = P(Y=K|X) = 1 - \sum_{k=1}^{K-1} P_k(X)$ as the baseline case.

(3.18) $\log_e[P_k(X)/P_K(X)] = \alpha_k + \beta_k X \Leftrightarrow P_k(X) = \exp(\alpha_k + \beta_k X)/[1 + \sum_{k=1}^{K-1} \exp(\alpha_k + \beta_k X)]$ for $k = 1 \ldots K-1$ and $P_K(X) = 1 - \sum_{k=1}^{K-1} P_k(X)$ (the baseline thus assigned the “left-over” probability)

(3.19) $\beta_k X = \beta_{k,1} X_1 + \beta_{k,2} X_2 + \ldots + \beta_{k,M} X_M$

with classes $k = \{1, \ldots, K-1\}$, and $M$ explanatory variables $X = \{X_1, \ldots, X_M\}$, parameters $\beta = \{(\beta_{1,1}, \ldots, \beta_{1,M}), (\beta_{2,1}, \ldots, \beta_{2,M}), \ldots, (\beta_{K-1,1}, \ldots, \beta_{K-1,M})\}$, and constants $\alpha = \{\alpha_1, \ldots, \alpha_{K-1}\}$.
One-vs-rest

- **Reasoning**: Each outcome is contrasted with the undifferentiated bulk of the rest
  - In principle could be simultaneously fitted!
- \{ajatella vs. \neg ajatella\}
  - \sim \{ajatella vs. \{miettiä, pohtia, harkita\}, …
- Number of binary models: \(n(\text{outcomes})\)
- Variables (and odds) distinguish individual outcomes against all the rest lumped together \(\rightarrow\) highlight outcome-specific distinctive features
- Direct probability estimates:
  - \(P(\text{outcome})\) generated directly, BUT
  - Normalization of probabilities required, so that \(\Sigma P(\text{all outcomes})=1\)
One-vs-rest

(3.23) $P_k(X) = P(Y=k|X)$, with and $k=\{1, \ldots, K\}$, and $P_{-k}(X) = P(Y=-k|X) = 1-P_k(X) = 1-P(Y=k|X)$ as the opposite case, that is, the ‘rest’, so naturally $P_k(X) + P_{-k}(X) = 1$ for each binary model.

(3.24) $\log_e \{P_k(X)/[1-P_k(X)]\} = \alpha_k + \beta_k X \iff P_k(X)/P_{-k}(X) = \exp(\alpha_k + \beta_k X)$

(3.25) $\beta_k X = \beta_{k,1}X_1 + \beta_{k,2}X_2 + \ldots + \beta_{k,M}X_M$

with classes $k=\{1, \ldots, K\}$, and $M$ explanatory variables $X=\{X_1, \ldots, X_M\}$, parameters $\beta = \{\beta_{1,1}, \ldots, \beta_{1,M}, \beta_{2,1}, \ldots, \beta_{2,M}, \ldots, \beta_{K,1}, \ldots, \beta_{K,M}\}$, and constants $\alpha=\{\alpha_1, \ldots, \alpha_K\}$
One-vs-rest
Pairwise contrasts

- **Reasoning:** all outcomes are contrasted pairwise with each other
  - \{ajatella vs. miettiä\}, \{ajatella vs. pohtia\}, \{ajatella vs. harkita\}, \{miettiä vs. ajatella\}, \{miettiä vs. pohtia\}, …

- Number of binary models:
  - Round-robin: \(\frac{n(\text{outcomes}) \cdot (n(\text{outcomes})-1)]}{2}\)
  - Double round-robin: \(n(\text{outcomes}) \cdot (n(\text{outcomes})-1)]\)

- Variables and odds sensitive to pairwise differences, but overall may exaggerate these and be difficult to interpret if distinctions are contradictory
  - Overall verb-feature odds can only be approximated as a geometric average of the pairwise odds

- No direct/approximate probability estimates
Pairwise contrasts

(3.26) $P_{k_1/k_2}(X) = [P(Y=k_1|X) \mid Y=\{k_1, k_2\}]$, and $P_{k_2/k_1}(X)=1-P_{k_1/k_2}(X) = 1-[P(Y=k_1|X) \mid Y=\{k_1, k_2\}]$

(3.27) $\log_e[P_{k_1/k_2}(X) \mid Y=\{k_1, k_2\}] = \alpha_{k_1/k_2} + \beta_{k_1/k_2}X$

(3.28) $\beta_{k_1/k_2}X = \beta_{k_1/k_2,1}X_1 + \beta_{k_1/k_2,2}X_2 + \ldots + \beta_{k_1/k_2,M}X_M$

(3.29) $\beta_{k_1,m} = (\beta_{k_1/k_2,m} + \beta_{k_1/k_3,m} + \ldots + \beta_{k_1/k,M})/(K-1)$, since the geometric average of the binary log-odds is $[e^{\beta(1)} \cdot e^{\beta(2)} \cdot \ldots \cdot e^{\beta(K-1)}]^{1/(K-1)} = e^{\beta(1) + \beta(2) + \ldots + \beta(K-1)/(K-1)}$

(3.30) $P_{k_1}(X) \approx \{n[P_{k_1/k_2}(X) > 0.5] + n[P_{k_2/k_1}(X) \leq 0.5]\}/[K \cdot (K-1)]$; N.B. $0 \leq P_{k_1}(X) \leq 0.5$

with classes $k_1=\{1, \ldots, K\}$, and $k_2=\{1, \ldots, K\}$, with $k_1 \neq k_2$, and $M$ explanatory variables $X=\{X_1, \ldots, X_M\}$, parameters $\beta = \{(\beta_{1/2,1}, \ldots, \beta_{1,M}), \ldots, (\beta_{1/K,1}, \ldots, \beta_{1/K,M}), (\beta_{2/1,1}, \ldots, \beta_{2/1,M}), \ldots, (\beta_{K/1,1}, \ldots, \beta_{K/1,M}), \ldots, (\beta_{K/K-1,1}, \ldots, \beta_{K/K-1,M})\}$, and constants $\alpha = \{\alpha_{k_1/k_2}, \alpha_{k_1/k_3}, \ldots, \alpha_{K/K-2}, \alpha_{K/K-1}\}$
Baseline vs. One-vs-rest vs. Pairwise contrasts

<table>
<thead>
<tr>
<th>Feature/Verb</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z_ANL_SG12</td>
<td>2</td>
<td>1/5</td>
<td>1/1</td>
</tr>
<tr>
<td>Z_ANL_SG3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Z_ANL_PL12</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>SX_AGESEM_INDIVIDUAL</td>
<td>1/1</td>
<td>1/1</td>
<td>1/2</td>
</tr>
<tr>
<td>SX_AGESEM_GROUP</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature/Verb</th>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z_ANL_SG12</td>
<td>(1)</td>
<td>2</td>
<td>1/6</td>
<td>(1/1)</td>
</tr>
<tr>
<td>Z_ANL_SG3</td>
<td>1/2</td>
<td>2</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>Z_ANL_PL12</td>
<td>1/2</td>
<td>2</td>
<td>(1/1)</td>
<td>(1)</td>
</tr>
<tr>
<td>SX_AGESEM_INDIVIDUAL</td>
<td>1</td>
<td>(1)</td>
<td>(1/1)</td>
<td>1/2</td>
</tr>
<tr>
<td>SX_AGESEM_GROUP</td>
<td>1/3</td>
<td>1/2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lexemes/Contextual features</th>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ANL_SINGULAR-1</td>
<td>2</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>0 ANL_SINGULAR-3</td>
<td>+/-</td>
<td>-</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>0 ANL_PLURAL-1</td>
<td>2</td>
<td>-</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>SX_AGENT+SEM_INDIVIDUAL</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>SX_AGENT+SEM_GROUP</td>
<td>---</td>
<td>+/-</td>
<td>+++</td>
<td>+</td>
</tr>
</tbody>
</table>
Nested dichotomy

- **Reasoning:** Polytomous setting is partitioned into a successive set of dichotomies (Fox 1997)
  - Partitioning should be clearly naturally motivatable
- E.g. \{ajatella vs. {miettiä vs. {pohtia vs. harkita}}\}
- Number of binary models: \(n(\text{outcomes})-1\)
  - N.B. number of partitions: \(T(1)=1;\)
    \[T[n(\text{outcomes})]= 2 \cdot n(\text{outcomes}-3) \cdot T(n(\text{outcomes})-1)\]
- Overall variable odds can be generated as a product of the sequence of odds
- Direct probability estimates can be calculated exactly as a product of the sequence of probabilities in the appropriate partitions
  - No normalization is necessary
Nested dichotomy

Consider e.g. the partition

\{ajatella vs. \{miettiä vs. \{pohtia vs. harkita\}\}\}

• The probability of the outcome \(Y=harkita\) for some given context and features (represented as \(X\)) is thus

\[P_{\{h\}|\{a,m,p\}}(Y=harkita|X)\]

\[P_{\{m,p,h\}|\{a\}}(Y=\{miettiä, pohtia, harkita\}|X)\]

\[\cdot P_{\{p,h\}|\{m\}}(Y=\{pohtia, harkita\}|X)\]

\[\cdot P_{\{h\}|\{p\}}(Y=\{harkita\}|X)\]
Ensemble of nested dichotomies

- **Reasoning:** Sample a set of partitions, when no obviously natural partitioning of the outcomes exists, and average over the results (Frank & Kramer 2004)
  - All partitions are considered equally likely, and may each represent fault-lines among the outcomes specific to one or more among the variables
  - 20 randomly sampled partitions sufficient

- Number of binary models: \(20 \cdot [n(\text{outcomes})-1]\)

- Overall variable odds may be approximated as an average of the aggregate odds of the constituent partitioned models; the same applies for outcome-specific probability estimates
## Summary overview – heuristics for polytomous logistic regression

<table>
<thead>
<tr>
<th>Heuristic/characteristics</th>
<th>Multinomial (baseline category)</th>
<th>One-vs-rest</th>
<th>Pairwise</th>
<th>Nested dichotomy</th>
<th>Ensemble of nested dichotomies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of constituent binary models</td>
<td>$n_{\text{lex}}^{-1}$</td>
<td>$n_{\text{lex}}$</td>
<td>$n_{\text{lex}}$, $(n_{\text{lex}}^{-1})/2$ (round-robin)</td>
<td>$n_{\text{lex}}^{-1}$</td>
<td>~20 partitions (each with $n_{\text{lex}}^{-1}$)</td>
</tr>
<tr>
<td>Lexeme-specific odds-ratios for feature variables</td>
<td>No (Every lexeme against the baseline)</td>
<td>Yes (Every lexeme against the rest)</td>
<td>No (Approximation by geometric averages of binary odds-ratios)</td>
<td>Yes (Products of binary odds-ratios)</td>
<td>Yes (Averages of products of binary odds-ratios)</td>
</tr>
<tr>
<td>Probability estimates for lexemes (i.e., outcomes)</td>
<td>Direct</td>
<td>Direct $P_{\text{lex/rest}}(X)$</td>
<td>No</td>
<td>Direct (Product of probabilities at nodes in partition tree)</td>
<td>Direct (Average of products of probabilities at nodes in partition tree)</td>
</tr>
<tr>
<td>Selection of lexeme in prediction</td>
<td>Probability-based $\arg_{\text{lex}} \max(P_{\text{lex}}</td>
<td>X)$</td>
<td>Probability-based $\arg_{\text{lex}} \max(P_{\text{lex}}</td>
<td>X)$</td>
<td>Voting $\arg_{\text{lex}} \max {n[P_{\text{lex1}}(X) &gt; 0.5] + n[P_{\text{lex2}}(X) &lt; 0.5]}$</td>
</tr>
<tr>
<td>Other</td>
<td>Necessity of baseline category</td>
<td>May not discover pairwise distinctions</td>
<td>May exaggerate pairwise distinctions, and the behavior with contradictory distinctions is problematic</td>
<td>Selection of single appropriate partition may be difficult or impossible</td>
<td>-</td>
</tr>
</tbody>
</table>
Comparisons of heuristics – model fit

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>$R_L^2$</th>
<th>Recall (%)</th>
<th>$\lambda_{\text{prediction}}$</th>
<th>$\tau_{\text{classification}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-vs-rest</td>
<td>0.313</td>
<td>64.60</td>
<td>0.370</td>
<td>0.490</td>
</tr>
<tr>
<td>pairwise</td>
<td>NA</td>
<td>64.63</td>
<td>0.370</td>
<td>0.490</td>
</tr>
<tr>
<td>(simultaneous) multinomial</td>
<td>0.316</td>
<td>64.89</td>
<td>0.375</td>
<td>0.494</td>
</tr>
<tr>
<td>ensemble of nested dichotomies (END)</td>
<td>0.315</td>
<td>64.78</td>
<td>0.373</td>
<td>0.493</td>
</tr>
<tr>
<td>“best” nested dichotomies: {A, {H, {M, P}}} and {P, {A, {M, H}}}</td>
<td>NA</td>
<td>64.66</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>“worst” nested dichotomy: {{A, P}, {M, H}}</td>
<td>NA</td>
<td>63.66</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
## Comparisons of heuristics – model fit

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>$R_L^2$</th>
<th>Recall (%)</th>
<th>$\lambda_{\text{prediction}}$</th>
<th>$\tau_{\text{classification}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-vs-rest</td>
<td>0.287 (0.264, 0.300)</td>
<td>63.80 (63.07, 64.51)</td>
<td>0.355 (0.343, 0.368)</td>
<td>0.479 (0.468, 0.489)</td>
</tr>
<tr>
<td>pairwise</td>
<td>NA</td>
<td>63.79 (62.87, 64.57)</td>
<td>0.355 (0.339, 0.369)</td>
<td>0.478 (0.465, 0.490)</td>
</tr>
<tr>
<td>(simultaneous) multinomial</td>
<td>0.292 (0.276, 0.302)</td>
<td>63.78 (62.96, 64.51)</td>
<td>0.355 (0.340, 0.368)</td>
<td>0.478 (0.466, 0.489)</td>
</tr>
<tr>
<td>ensemble of nested dichotomies (END)</td>
<td>0.294 (0.277, 0.305)</td>
<td>63.89 (63.10, 64.63)</td>
<td>0.357 (0.343, 0.370)</td>
<td>0.480 (0.468, 0.490)</td>
</tr>
<tr>
<td>“best” nested dichotomy: {A, {H, {M, P}}}</td>
<td>NA</td>
<td>63.65 (62.87, 64.37)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>“worst” nested dichotomy: {A, {P, {M, H}}}</td>
<td>NA</td>
<td>63.01 (61.93, 63.84)</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
# Comparisons of heuristics – overlap of outcome selections

<table>
<thead>
<tr>
<th>Heuristics</th>
<th>pairwise</th>
<th>multinomial (simultaneous)</th>
<th>ensemble of nested dichotomies</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-vs-rest</td>
<td>3279 (96.3%)</td>
<td>3325 (97.7%)</td>
<td>3360 (98.7%)</td>
</tr>
<tr>
<td>pairwise</td>
<td>-</td>
<td>3313 (97.3%)</td>
<td>3312 (97.3%)</td>
</tr>
<tr>
<td>multinomial (simultaneous)</td>
<td>-</td>
<td>-</td>
<td>3344 (98.2%)</td>
</tr>
</tbody>
</table>
Results – overall probabilities estimated by the full model

Figure 1. Probabilities of lexemes per each context.
Results – probabilities

- only 258 (7.6%) instances for which $P_{\text{max}}(L|C) > 0.90$
- as many as 764 (22.4%) of the minimum estimated probabilities per instance are practically nil with $P_{\text{min}}(L|C) < 0.01$
- the other way around, for 2640 (77.6%) instances the minimum estimated probability $P_{\text{min}}(L|C) \geq 0.01$
  - i.e. representing an expected possibility of occurrence at least once every hundred times or even more often in a similar context.
Model fit revisited – Proportions vs. Probabilities

- Another frequent feature combination in data:
  - $n\{Z_{ANL\_IND}, Z_{ANL\_THIRD}, SX_{AGE\_SEM\_GROUP}, SX_{PAT\_SEM\_ACTIVITY}\}=17$

- Observed frequencies

<table>
<thead>
<tr>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

- Observed proportions

<table>
<thead>
<tr>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.06</td>
<td>0.24</td>
<td>0.71</td>
</tr>
</tbody>
</table>

- Estimated probabilities

<table>
<thead>
<tr>
<th>ajatella</th>
<th>miettiä</th>
<th>pohtia</th>
<th>harkita</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>0.06</td>
<td>0.41</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Still another frequent feature combination in data:
  • \( n\{Z\_PHR\_CLAUSE, SX\_PAT\_SEM\_ABSTRACTION\} = 31 \)

- **Observed frequencies**
  
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ajatella</td>
<td>miettiä</td>
<td>pohtia</td>
<td>harkita</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

- **Observed proportions**
  
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ajatella</td>
<td>miettiä</td>
<td>pohtia</td>
<td>harkita</td>
</tr>
<tr>
<td>0.29</td>
<td>0.13</td>
<td>0.32</td>
<td>0.26</td>
</tr>
</tbody>
</table>

- **Estimated probabilities**
  
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ajatella</td>
<td>miettiä</td>
<td>pohtia</td>
<td>harkita</td>
</tr>
<tr>
<td>0.33</td>
<td>0.12</td>
<td>0.39</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Exemplary contexts of usage – *ajatella*

<table>
<thead>
<tr>
<th>A:#1 (7/2)</th>
<th>Miten MANNER+GENERIC <em>ajattelit</em> INDICATIVE +SECOND, COVERT, AGENT+INDIVIDUAL erotapATIENT+INFINITIVE … jostain SAKn kannattajasta? [sfnet] [3066/politiikka_9967] ‘How did you think to differ at all from some dense supporter of class-thinking in SAK?’</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(<em>ajatella</em></td>
<td>Context)=1</td>
</tr>
<tr>
<td>P(<em>pohtia</em></td>
<td>Context)=0</td>
</tr>
<tr>
<td>M:#2 (7/1)</td>
<td>P(ajatella</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------</td>
</tr>
</tbody>
</table>

**Vilkaise**<sup>CO-ORDINATED VERB(MENTAL)</sup> **joskus**<sup>FREQUENCY(SOMETIMES)</sup> **valtuuston esityslistaa ja mieti**<sup>IMPERATIVE+, SECOND, COVERT, AGENT+INDIVIDUAL</sup> **monestako**<sup>PATIENT</sup> **+INDIRECT_QUESTION** **asiasta sinulla on jotain tietoa.** [sfnet]

‘Glance sometimes at the agenda for the council and think on how many issues you have some information.’
Exemplary contexts – *pohtia*

| P(#1 (6/3) P(ajatella|Context)=0.036 | Suomessa *LOCATION(+LOCATION)* kansalaisjärjestöt *AGENT+GROUP* pohtivat *INDICATIVE+THIRD+PLURAL …* auttamisen periaatteita *PATIENT+NOTION …* eettisessä neuvottelukunnassa *LOCATION(+GROUP).* [1259/ hs95_10437] |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| P(miettiäContext)=0.071 | ‘In Finland civic organizations are pondering the principles of novel forms of assistance (e.g. the identification of an A-subscriber) in the so-called ethical advisory board of telephone assistance.’ |
| P(pohtia|Context)=0.852 | |
| P(harkita|Context)=0.041 | |
Exemplary contexts – *harkita*

| H:#1 (7/2) | Monen puoluetoverinkin mielestä META … Kauko Juhantalon AGENT+INDIVIDUAL olisi CONDITIONAL+THIRD pitänyt VERB_CHAIN +NECESSITY *harkita* tarkemmin MANNER +POSITIVE(<THOROUGH) ehdokkuuttaan. [275/hs95_2077] ‘In the opinion of many fellow party members, for instance Kauko Juhantalo should have considered more carefully his candidacy.’ |
|------------|--------------------------------------------------|----------------------------------------------|
| P(ajatella|Context)=0.025 | *Monen puoluetoverinkin mielestä META … Kauko Juhantalon AGENT+INDIVIDUAL olisi CONDITIONAL+THIRD pitänyt VERB_CHAIN +NECESSITY* |
| P(miettiäContext)=0.115 | *harkita* tarkemmin MANNER +POSITIVE(<THOROUGH) ehdokkuuttaan. [275/hs95_2077] ‘In the opinion of many fellow party members, for instance Kauko Juhantalo should have considered more carefully his candidacy.’ |
| P(pohtia|Context)=0.135 | |
| P(harkita|Context)=**0.725** | |

*In the opinion of many fellow party members, for instance Kauko Juhantalo should have considered more carefully his candidacy.’*
Variation – “wrong” choice

| H:#2 (8/2) | P(ajatella|Context)=0.025 |
|------------|---------------------------|
| P(miettiä|Context)=0.125 |
| P(pohtia|Context)=0.125 |
| P(harkita|Context)=0.725 |

\textbf{Tarkastusviraston mielestä} META
\textbf{tätä ehdotusta} PATIENT+ACTIVITY
\textbf{oli} CONDITIONAL+THIRD, COVERT
\textbf{syytä} VERB_CHAIN+NECESSITY
\textbf{pohtia tarkemmin} MANNER+POSITIVE.
[766/hs95_7542]

‘In the opinion of the Revision Office there is reason to \textbf{ponder} this proposal more thoroughly.’
<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ajatella</td>
<td>0.301</td>
</tr>
<tr>
<td>miettiä</td>
<td>0.272</td>
</tr>
<tr>
<td>pohtia</td>
<td>0.215</td>
</tr>
<tr>
<td>harkita</td>
<td>0.212</td>
</tr>
</tbody>
</table>

Variation – equiprobable choice

Aluksi harvemmin, mutta myöhemmin tyttö alko \(\text{viettää öitä } T:n \text{ luona ja vuoden tapailun päätteeksi} \]

\( P_{\text{AGENT+INDIVIDUAL}} \text{ sanoi, että voisi CONDITIONAL} +\text{THIRD, VERB-CHAIN+POSSIBILITY, COVERT} \)

\( \text{ajatella asiaa PATIENT+ABSTRACTION(<NOTION)} \)

\( \text{vakavamminkin MANNER+POSITIVE. [sfnet] [50/ihmissuhteet_8319]} \)

‘... P said that [he] could think about the matter more seriously [perhaps]’
Syonymy – or not?

- Aluksi harvemmin, mutta myöhemmin tyttö alko vietä öitä T:n luona ja vuoden tapailun päätteeksi P sanoi, että voisi ajatella asiaa vakavamminkin.
  - Possibility to have an attitude/opinion concerning the 'issue' (asia)
  ... P sanoi, että voisi miettiä asiaa vakavamminkin.
  - Actually give some occasional thought to the 'issue', without any expression of its duration, intensity
  ... P sanoi, että voisi pohtia asiaa vakavamminkin.
  - Give the 'issue' serious, considerable and lengthy consideration
  ... P sanoi, että voisi harkita asiaa vakavamminkin.
  - Consider the 'issue' with respect to making a decision one way or another concerning it
- None of these can be resolved on the basis of the immediate sentence context alone
- Might be deducible from prior passages in the text or extralinguistic knowledge about the context and/or the participants in the linguistic exchange
Results - discussion

- The recall rate seems to reach a ceiling at ~65%, and appears indifferent to whether some individual group of variables is left out
  - Do we yet lack some necessary variables or variable types?
  - Are some of the characteristics embedded in the synonymous lexeme itself, and not manifest – nor expressable – in any overt way in the immediate context (though possibly in the entire text or overall extralinguistic context)
  - Does this level represent the maximum that can be reached with the descriptive apparatus and associated variables of traditional grammatical analysis?
  - Might the remaining one-third represent to some extent cases of “true” synonymy and interchangeability?

- The results support Bresnan’s (2007) probabilistic view of the relationship between linguistic usage and the underlying linguistic system
  - Few choices are categorical, given the known context (feature cluster) that can be analytically grasped and identified
  - Rather, most contexts exhibit various degrees of variation as to their outcomes, resulting in proportionate choices on the long run
  - The question remains to what extent we are able to model this variation on the basis of current conventional linguistic theories

- These should be tested by comparing the predicted probabilities with selection in forced-choice experiments as well as acceptability ratings
The End

- Thank you!
- Questions, comments, suggestions?!?
Results (statistical)

- Measures of overall fit (Menard 1995)
  - recall rate = 65.6%
  - $R_L^2 = 0.325$
  - $\lambda_{\text{prediction}} = 0.387$
  - $\tau_{\text{classification}} = 0.504$

- Measures of model validation
  - 1000 repetitions of training the model with a simple bootstrap resample and then testing the model against the entire data, on the basis of which we finally calculate a mean and the 95% Confidence Intervals of the model statistics
  - recall rate = 63.8% (63.07-64.51%)
  - $R_L^2(\text{TEACH}) = 0.325$ (0.307, 0.342)
  - $R_L^2(\text{TEST}) = 0.287$ (0.264, 0.300)
  - $\lambda_{\text{prediction}} = 0.355$ (0.343, 0.368)
  - $\tau_{\text{classification}} = 0.479$ (0.468, 0.489)

- Compare these with the 58-59% recall rate reported by Arppe (2006) using only semantic classifications of nominals (WordNet)
### Results - comparison of models with different sets of feature categories I

<table>
<thead>
<tr>
<th>Feature set composition</th>
<th>Recall (%)</th>
<th>$R_L^2$</th>
<th>$\lambda_{\text{prediction}}$</th>
<th>$\tau_{\text{classification}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb-chain general morphological features (10) as well as those node-specific features which are not subsumed by the verb-chain general ones (17)</td>
<td>47.71</td>
<td>0.100</td>
<td>0.069</td>
<td>0.247</td>
</tr>
<tr>
<td>Syntactic argument types, <em>without</em> their semantic and structural classifications</td>
<td>50.18</td>
<td>0.098</td>
<td>0.113</td>
<td>0.282</td>
</tr>
<tr>
<td>Extralinguistic features alone (2)</td>
<td>47.21</td>
<td>0.057</td>
<td>0.060</td>
<td>0.240</td>
</tr>
</tbody>
</table>
## Results - comparison of models with different sets of feature categories II

<table>
<thead>
<tr>
<th>Feature set composition</th>
<th>Recall (%)</th>
<th>$R_L^2$</th>
<th>$\lambda_{prediction}$</th>
<th>$\tau_{classification}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model with verb-chain general morphological features (10) and their semantic classifications (6) together with syntactic argument types alone (10) or their selected or collapsed subtypes (20)</td>
<td>64.60</td>
<td>0.313</td>
<td>0.370</td>
<td>0.490</td>
</tr>
<tr>
<td>Full model with verb-chain general morphological features (10) and their semantic classifications (6) together with syntactic argument types alone (10) or their subtypes (20) as well as extra-linguistic features (2)</td>
<td>65.57</td>
<td>0.325</td>
<td>0.387</td>
<td>0.504</td>
</tr>
</tbody>
</table>
## Results - relative importance of feature categories in final model

<table>
<thead>
<tr>
<th>Feature variable category</th>
<th>Mean odds in favor</th>
<th>Mean odds against</th>
<th>Mean aggregate odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb chain morphology</td>
<td>2.02 (1.48)</td>
<td>0.52<del>1:1.91 (0.70</del>1:1.44)</td>
<td>1.96 (1.46)</td>
</tr>
<tr>
<td>Verb chain semantics</td>
<td>2.66 (1.62)</td>
<td>0.24<del>1:4.24 (0.12</del>1:8.59)</td>
<td>3.17 (3.48)</td>
</tr>
<tr>
<td>Syntactic argument types (alone)</td>
<td>2.69 (1.84)</td>
<td>0.32<del>1:3.14 (0.47</del>1:2.11)</td>
<td>2.92 (1.99)</td>
</tr>
<tr>
<td>Syntax arguments + semantic/structural subtypes</td>
<td>3.71 (2.57)</td>
<td>0.21<del>1:4.70 (0.06</del>1:18)</td>
<td>4.13 (7.89)</td>
</tr>
<tr>
<td>Extralinguistic features</td>
<td>1.68 (1.68)</td>
<td>0.47<del>1:2.13 (0.56</del>1:1.80)</td>
<td>1.86 (1.74)</td>
</tr>
</tbody>
</table>
### Results (linguistic) - Odds(lexeme <-- feature)

<table>
<thead>
<tr>
<th>Lexeme/Features</th>
<th>Strongest odds in favor of the lexeme</th>
<th>Strongest odds against the lexeme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ajatella</strong></td>
<td>SX_MAN.SEM_GENERIC (23)</td>
<td>SX_PAT.DIRECT_QUOTE (0.013~1:75)</td>
</tr>
<tr>
<td></td>
<td>SX_MAN.SEM_AGREEMENT (16)</td>
<td>SX_PAT.INDIRECT_QUESTION (0.07~1:14)</td>
</tr>
<tr>
<td></td>
<td>SX_VCH.SEM_ACCIDENTAL (5.6)</td>
<td>SX_PAT.SEM_COMMUNICATION (0.1~1:9.6)</td>
</tr>
<tr>
<td></td>
<td>SX_PAT.INFINITIVE (5.3)</td>
<td>SX_PAT.SEM_ACTIVITY (0.14~1:7.1)</td>
</tr>
<tr>
<td></td>
<td>SX_PAT.PARTICIPLE (5.3)</td>
<td></td>
</tr>
<tr>
<td><strong>miettiä</strong></td>
<td>SX_PAT.INDIRECT_QUESTION (4.2)</td>
<td>SX_MAN.SEM_AGREEMENT (0.07~1:14)</td>
</tr>
<tr>
<td></td>
<td>SX_DUR (3.4)</td>
<td>SX_MAN.SEM_GENERIC (0.15~1:6.8)</td>
</tr>
<tr>
<td></td>
<td>SX_PAT.DIRECT_QUOTE (3)</td>
<td>SX_MAN.SEM_FRAME (0.28~1:3.6)</td>
</tr>
<tr>
<td></td>
<td>SX_PAT.SEM_COMMUNICATION (2.8)</td>
<td>SX_AGE.SEM_GROUP (0.52~1:1.9)</td>
</tr>
<tr>
<td></td>
<td>SX_QUA (2.6)</td>
<td>SX_LX_että_CS.SX_PAT (0.52~1:1.9)</td>
</tr>
<tr>
<td>Lexeme/Features</td>
<td>Strongest odds in favor of the lexeme</td>
<td>Strongest odds against the lexeme</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td><strong>pohtia</strong></td>
<td>SX_PAT.DIRECT_QUOTE (8.1)</td>
<td>SX_MAN.SEM_AGREEMENT (0.22~1:4.5)</td>
</tr>
<tr>
<td></td>
<td>SX_AGE.SEM_GROUP (4.2)</td>
<td>SX_MAN.SEM_NEGATIVE (0.22~1:4.6)</td>
</tr>
<tr>
<td></td>
<td>SX_PAT.SEM_ABSTRACTION (4.1)</td>
<td>SX_SOU (0.29~1:3.5)</td>
</tr>
<tr>
<td></td>
<td>SX_LOC (3.7)</td>
<td>Z_ANL_FIRST (0.29~1:3.5)</td>
</tr>
<tr>
<td></td>
<td>SX_PAT.SEM_COMMUNICATION (3)</td>
<td>SX_PAT.SEM_INDIV..._GROUP (0.3~1:3.4)</td>
</tr>
<tr>
<td><strong>harkita</strong></td>
<td>SX_PAT.SEM_ACTIVITY (9)</td>
<td>SX_SOU (0.13~1:7.5)</td>
</tr>
<tr>
<td></td>
<td>SX_CND (2.9)</td>
<td>SX_VCH.SEM_TEMPORAL (0.15~1:6.5)</td>
</tr>
<tr>
<td></td>
<td>Z_ANL_KOND (2.3)</td>
<td>SX_GOA (0.21~1:4.7)</td>
</tr>
<tr>
<td></td>
<td>SX_MAN.SEM_POSITIVE (1.8)</td>
<td>SX_LX_että_CS.SX_PAT (0.25~1:4.5)</td>
</tr>
<tr>
<td></td>
<td>SX_META (1.6)</td>
<td>SX_MAN.SEM_FRAME (0.27~1:3.8)</td>
</tr>
</tbody>
</table>
## Results (linguistic) - Odds (feature --> lexeme)

<table>
<thead>
<tr>
<th>Contextual Feature</th>
<th>Lexemes with strong odds in favor</th>
<th>Lexemes with neutral odds</th>
<th>Lexemes with strong odds against</th>
</tr>
</thead>
<tbody>
<tr>
<td>SX_AGE.SEM_INDIVIDUAL</td>
<td>-</td>
<td>pohtia (1.6), miettiä (0.98), ajatella (0.85), harkita (0.69)</td>
<td>-</td>
</tr>
<tr>
<td>SX_AGE.SEM_GROUP</td>
<td>pohtia (4.2)</td>
<td>harkita (1.1)</td>
<td>miettiä (0.52), ajatella (0.2)</td>
</tr>
<tr>
<td>Z_ANL_FIRST</td>
<td>-</td>
<td>harkita (1.9), miettiä (1.8), ajatella (0.86)</td>
<td>pohtia (0.29)</td>
</tr>
<tr>
<td>Z_ANL_SECOND</td>
<td>miettiä (2.4)</td>
<td>ajatella (0.69), harkita (0.68)</td>
<td>pohtia (0.42)</td>
</tr>
<tr>
<td>Z_ANL_THIRD</td>
<td>-</td>
<td>harkita (1.6), miettiä (1.3), pohtia (0.99), ajatella (0.63)</td>
<td>-</td>
</tr>
<tr>
<td>Z_ANL_PLUR</td>
<td>pohtia (1.6)</td>
<td>harkita (1.2), ajatella (1.1)</td>
<td>miettiä (0.59)</td>
</tr>
<tr>
<td>Z_ANL_PASS</td>
<td>pohtia (1.9)</td>
<td>harkita (1.1), miettiä (0.89), ajatella (0.63)</td>
<td>-</td>
</tr>
<tr>
<td>Z_ANL_COVERT</td>
<td>-</td>
<td>miettiä (1.2), ajatella (1.1), harkita (0.79), pohtia (0.77)</td>
<td>- 59</td>
</tr>
<tr>
<td>Rank</td>
<td>A</td>
<td>M</td>
<td>P</td>
</tr>
<tr>
<td>------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
| A:#1 | 1 | 0 | 0 | 0 | *Mitenniferajattelit* indicative
  +second, covert, agent+individual
  +erota+infinite mitenkään
  jostain SAKn umpimielisistä luokka-
  ajattelun kannattajasta? [3066/
politiikka_9967]
  ‘How did you **think** to differ at all
  from some dense supporter of class-
  thinking in SAK?’ |
### Results (linguistic) - Estimated probabilities - dispersion

<table>
<thead>
<tr>
<th>Rank</th>
<th>A</th>
<th>M</th>
<th>P</th>
<th>H</th>
<th>Sentences</th>
</tr>
</thead>
</table>
| M:#2  | 0.018| 0.878| 0.084| 0.02 | **Vilkaise**<sub>CO-ORDINATED_VERB(MENTAL)</sub> **joskus**<sub>FREQUENCY(SOMETIMES)</sub> **valtuuston esityslistaa ja mieti**<sub>IMPERATIVE(SECOND, COVERT, AGENT+INDIVIDUAL monestako PATIENT +INDIRECT_QUESTION asiasta sinulla on jotain tietoa. [2815/politiikka_728]**

‘Glance sometimes at the agenda for the council and **think** on how many issues you have some information.’
Results (linguistic) - Estimated probabilities – “wrong” selection

<table>
<thead>
<tr>
<th>Rank</th>
<th>A</th>
<th>M</th>
<th>P</th>
<th>H</th>
<th>Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>H:#2 (8/2)</td>
<td>0.025</td>
<td>0.125</td>
<td><strong>0.125</strong></td>
<td><strong>0.725</strong></td>
<td>&quot;Tarkastusviraston mielestä tätä ehdotusta Patient+Activity olisi Conditional+Third syytä Verb_Chain+Necessity pohdia tarkemmin Manner+Positive. [766/hs95_7542] ‘In the opinion of the revision office there is reason to ponder this proposal more thoroughly.’&quot;</td>
</tr>
</tbody>
</table>
Results (linguistic) -
Estimated probabilities –
P(all)~0.25 -> synonymy?

<table>
<thead>
<tr>
<th>$\sigma(P)$</th>
<th>A</th>
<th>M</th>
<th>P</th>
<th>H</th>
<th>Sentences</th>
</tr>
</thead>
</table>
| 0.044      | 0.301 | 0.272 | 0.215 | 0.212 | *Aluksi harvemmin, mutta myöhemmin tyttö alkoiviettää öitä T:n luona ja vuoden tapailun pääteeksi* $P_{\text{AGENT+INDIVIDUAL}}$ sanoi, että voisi $P_{\text{CONDITIONAL}}$ +THIRD,VERB-CHAIN+POSSIBILITY,COVERT *ajatella asiaa* $P_{\text{PATIENT+NOTION}}$ vakavamminkin $P_{\text{MANNER+POSITIVE}}$ [50/ihmissuhteet_8319] ‘... P said that [she] could think about the matter more seriously [perhaps]’
Synonymy – or not?

- Aluksi harvemmin, mutta myöhemmin tyttö alkoiviettää öitä T:n luona ja vuoden tapailun päätteeksi P sanoi, että voisi ajatella asiaa vakavamminkin.
  - Possibility to have an attitude/opinion concerning the 'issue' (asia)
- ... P sanoi, että voisi miettiä asiaa vakavamminkin.
  - Actually give some occasional thought to the 'issue', without any expression of its duration, intensity
- ... P sanoi, että voisi pohtia asiaa vakavamminkin.
  - Give the 'issue' serious, considerable and lengthy consideration
- ... P sanoi, että voisi harkita asiaa vakavamminkin.
  - Consider the 'issue' with respect to making a decision one way or another concerning it
- None of these can be resolved on the basis of the immediate sentence context alone
- Might be deducible from prior passages in the text or extralinguistic knowledge about the context and/or the participants in the linguistic exchange
Results - overall probabilities estimated by the full model

Figure 1. Probabilities of lexemes per each context.
Results - overall probabilities estimated per lexeme by the full model
Results - overall probabilities estimated per lexeme by the full model
Results - probabilities

- maximum probability per all instances and contexts
  - Mean as low as \( x(P_{\text{max}}[L|C]) = 0.636 \),
  - overall span of maximal values as broad as (0.28, 1.00)
  - 95% CI = (0.369, 0.966).

- second-highest probability estimates per instances
  - mean \( x(P_{\text{max}-1}[L|C]) = 0.244 \)
  - overall range of (0.000, 0.490)
  - 95% CI = (0.026, 0.415)

- third-highest probability estimates
  - mean \( x(P_{\text{max}-2}[L|C]) = 0.096 \)
  - overall range of (0.000, 0.307)
  - 95% CI = (0.000, 0.241)

- minimum probability estimates
  - clearly keep some distance from zero as their mean \( x(P_{\text{min}}[L|C]) = 0.043 \)
  - even though their overall range is (0.000, 0.212) as well as 95% CI = (0.000, 0.144)
Results – probabilities

- only 258 (7.6%) instances for which \( P_{max}(L|C) > 0.90 \)
- as many as 764 (22.4%) of the minimum estimated probabilities per instance are practically nil with \( P_{min}(L|C) < 0.01 \)
- the other way around, for 2640 (77.6%) instances the minimum estimated probability \( P_{min}(L|C) \geq 0.01 \)
  - i.e. representing an expected possibility of occurrence at least once every hundred times or even more often in a similar context.
Results - discussion

- The recall rate seems to reach a ceiling at ~65%, and appears indifferent to whether some individual group of variables is left out
  - Do we yet lack some necessary variables or variable types?
  - Are some of the characteristics embedded in the synonymous lexeme itself, and not manifest – nor expressable – in any overt way in the immediate context (though possibly in the entire text or overall extralinguistic context)
  - Does this level represent the maximum that can be reached with the descriptive apparatus and associated variables of traditional grammatical analysis?
  - Might the remaining one-third represent to some extent cases of “true” synonymy and interchangeability?

- The results support Bresnan’s (2007) probabilistic view of the relationship between linguistic usage and the underlying linguistic system
  - Few choices are categorical, given the known context (feature cluster) that can be analytically grasped and identified
  - Rather, most contexts exhibit various degrees of variation as to their outcomes, resulting in proportionate choices on the long run
  - The question remains to what extent we are able to model this variation on the basis of current conventional linguistic theories

- These should be tested by comparing the predicted probabilities with selection in forced-choice experiments as well as acceptability ratings
The End

- Thank you!
- Questions, comments, suggestions?!?