

# Probabilistic Semantic Frames

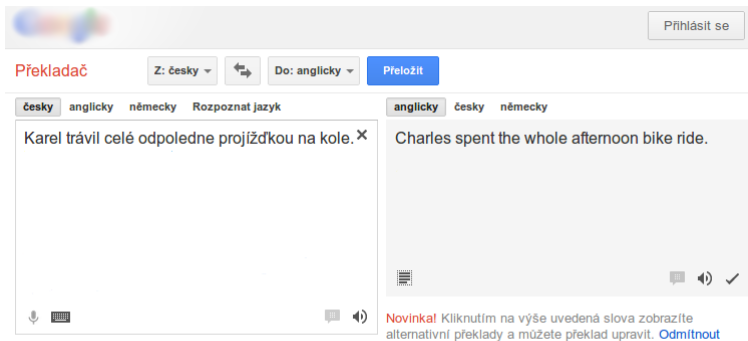
(IB030)

Jiří Materna

Centre for Natural Language Processing  
Faculty of informatics, Masaryk University  
Brno, Czech Republic

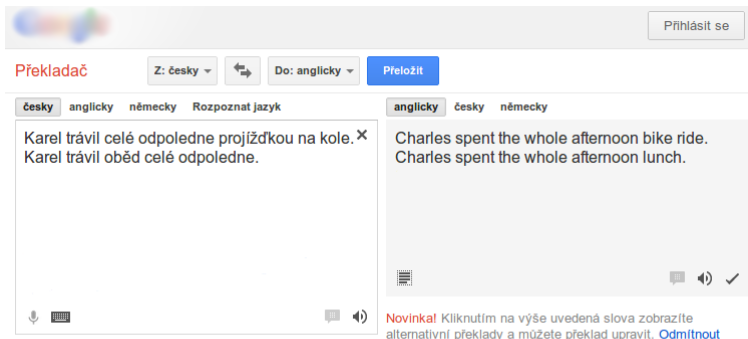
May 11, 2015

# Motivation I



The screenshot shows the Google Translate web interface. At the top right, there is a button labeled "Přihlásit se". Below it, the word "Překladač" is displayed in red. The interface includes two dropdown menus for source and target languages: "Z: česky" and "Do: anglicky", with a bidirectional arrow icon between them. A blue "Přeložit" button is positioned to the right. Below the language settings, there are two panels. The left panel, labeled "česky", contains the text "Karel trávil celé odpoledne projíždkou na kole. ✕". The right panel, labeled "anglicky", contains the translation "Charles spent the whole afternoon bike ride." Below the translation, there are icons for a dictionary, a speech bubble, a speaker icon, and a checkmark. At the bottom right, a red "Novinka!" notice states: "Kliknutím na výše uvedená slova zobrazíte alternativní překlady a můžete překlad upravit. [Odmítnout](#)".

# Motivation II



Přihlásit se

Překladač

Z: **česky** ↔ Do: **anglicky** Přeložit

**česky** anglicky německy Rozpoznat jazyk

Karel trávil celé odpoledne projíždkou na kole. ✕  
Karel trávil oběd celé odpoledne.

**anglicky** česky německy

Charles spent the whole afternoon bike ride.  
Charles spent the whole afternoon lunch.

Novinka! Kliknutím na výše uvedená slova zobrazíte alternativní překlady a můžete překlad upravit. [Odmítnout](#)

# Motivation III

Přihlásit se

Překladač Z: **česky** ↔ Do: **anglicky** Přeložit

**česky** anglicky německy Rozpoznat jazyk

Karel trávil celé odpoledne projížďkou na kole. X  
Karel trávil oběd celé odpoledne.  
Karel trávil krysy prudkým jedem.

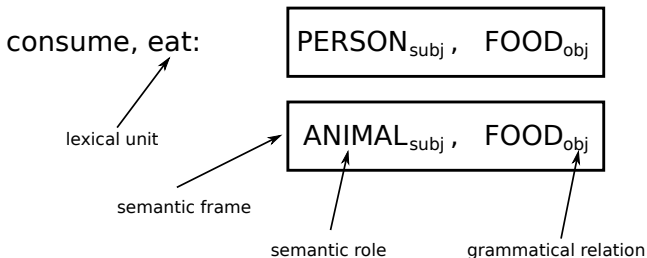
**anglicky** česky německy

Charles spent the whole afternoon bike ride.  
Charles spent the whole afternoon lunch.  
Charles spent sharp rat poison.

Novinka! Kliknutím na výše uvedená slova zobrazíte alternativní překlady a můžete překlad upravit. [Odmítnout](#)

# Semantic frames

- terminology adopted from Frame Semantics
- captures selectional preferences of grammatical relations



## Criticism of current DBs of SFs

- development of SFs is very time-consuming and expensive
- small coverage
- no statistical information
- subjectively biased

# Objectives

- semantic frames should be generated automatically from corpora
- algorithm must be unsupervised
- model should be probabilistic
- model should provide a possibility to estimate its parameters automatically
- set of semantic roles should be shared between different SFs

# LDA-Frames

- unsupervised method for discovering semantic frames
- generative model inspired by Latent Dirichlet Allocation
- language independent
- for each lexical unit a probability distribution over frames
- semantic roles represented as probability distributions over words
- need for a syntactically annotated corpus
- labels for roles must be assigned manually (if required)



# Web demo

EAT

	SUBJECT	OBJECT
	222	40
0.554086 frame 1166	0.794216 person	0.085888 food
	0.010335 people	0.046396 meal
	0.007963 one	0.01947 egg
	0.005797 man	0.01947 breakfast
	0.004342 who	0.01726 lunch
	0.003409 woman	0.016846 dinner
	0.002687 child	0.015189 fish
	0.002519 that	0.013256 meat
	0.002307 all	0.012289 potato
	0.002215 someone	0.012151 cake
	152	40
0.128011 frame 622	0.027104 bird	0.085888 food
	0.026926 dog	0.046396 meal
	0.023538 animal	0.01947 egg
	0.023181 fish	0.01947 breakfast
	0.016049 cat	0.01726 lunch
	0.014979 child	0.016846 dinner
	0.013374 people	0.015189 fish
	0.01266 prey	0.013256 meat
	0.011947 man	0.012289 potato
	0.011769 horse	0.012151 cake

## LDA-frames – training data

Set of frame realizations for each lexical unit

<b>Lexical unit</b>	<b>subject</b>	<b>object</b>	<b>frame</b>
eat	John Mike man	food pizza cake	(Person, Food)
	dog mouse	meat cheese	(Animal, Food)
drink	Jane Mike	coffee tee	(Person, Drink)
teach	teacher professor	student Mike	(Person, Person)
	Peter	dog	(Person, Animal)

Table : Example of grammatical relation realizations.

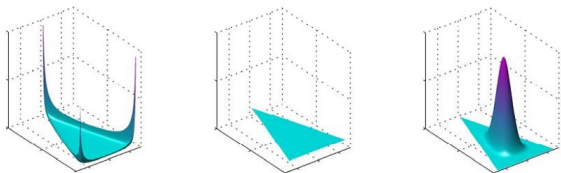
# Multinomial and Dirichlet distributions

Multinomial distribution:

$$P(\vec{m}|\vec{p}) = \frac{(\sum_{i=1}^k m_i)!}{m_1! m_2! \dots m_k!} \prod_{i=1}^k p_i^{m_i}, \quad (1)$$

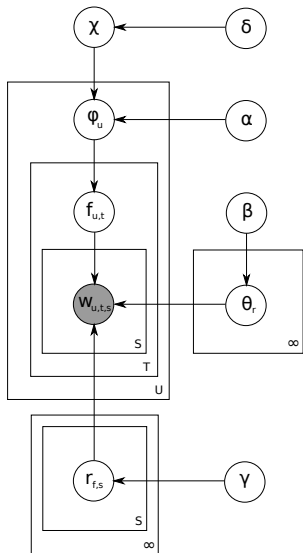
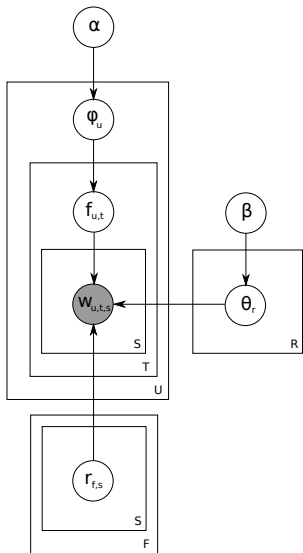
Dirichlet distribution:

$$\text{Dir}(\vec{p}|\vec{\alpha}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_k)} \prod_{i=1}^k p_i^{\alpha_i-1}, \quad (2)$$

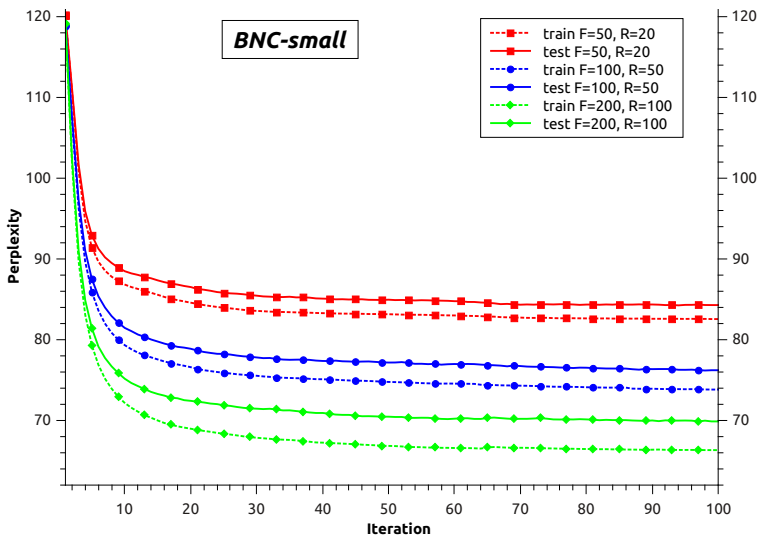


**Figure 2.5** Plots of the Dirichlet distribution over three variables, where the two horizontal axes are coordinates in the plane of the simplex and the vertical axis corresponds to the value of the density. Here  $\{\alpha_k\} = 0.1$  on the left plot,  $\{\alpha_k\} = 1$  in the centre plot, and  $\{\alpha_k\} = 10$  in the right plot.

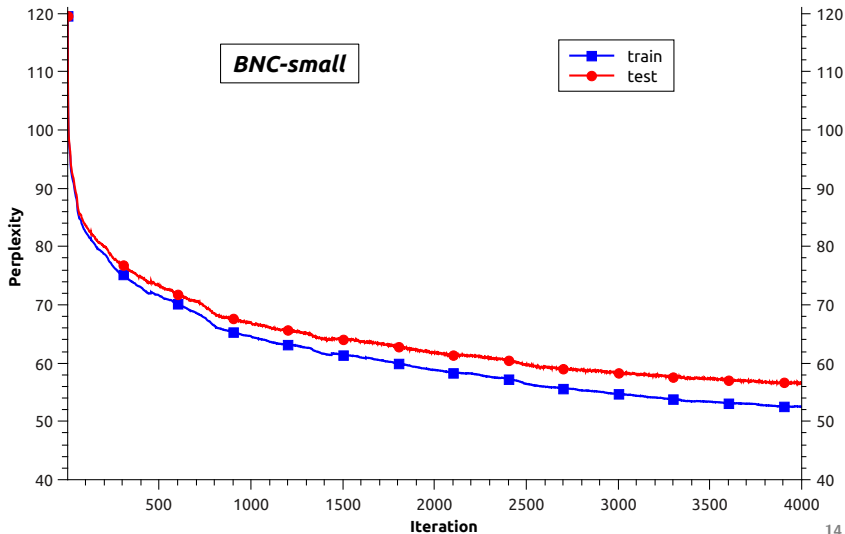
# LDA-Frames models



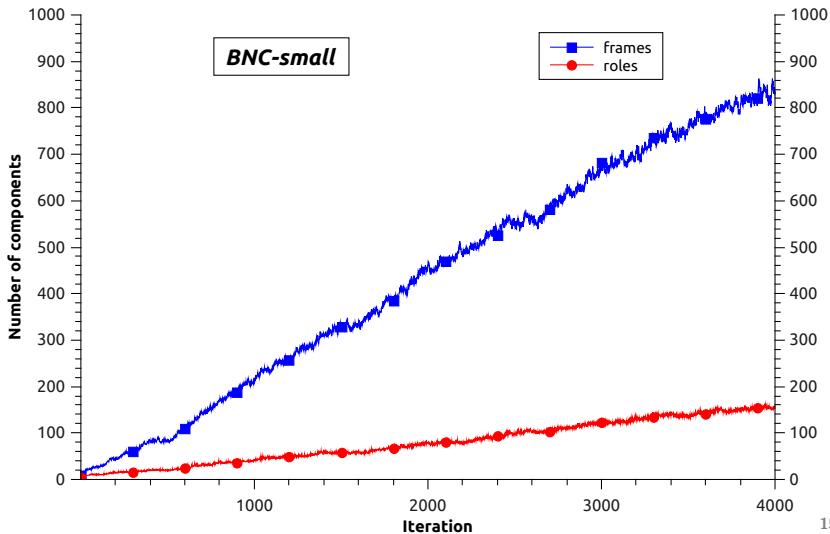
# LDA-Frames perplexity



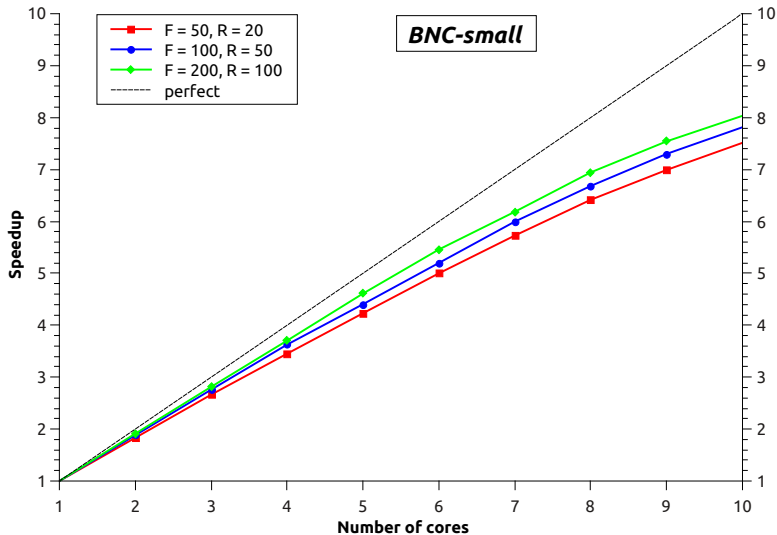
# Non-Parametric LDA-Frames perplexity



# Non-Parametric LDA-Frames parameters



# Scalability of Parallel LDA-Frames





## Reconstruction abilities I

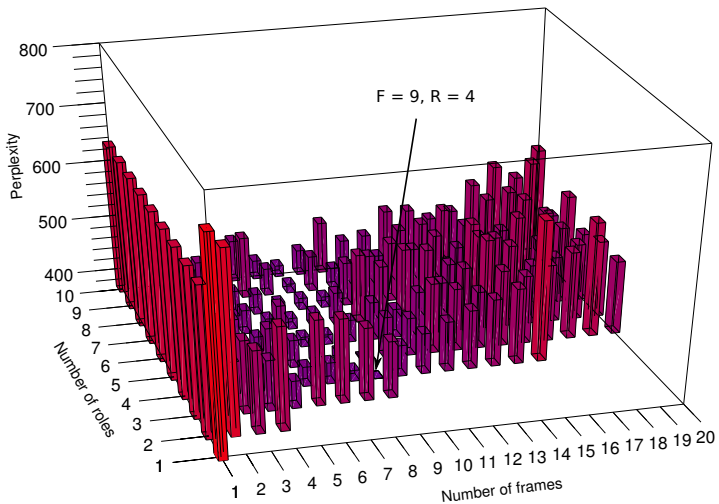
<b>semantic roles</b>	<b>realizations</b>
Animal	dog, cat, mouse, fish, chicken, jenny, rabbit
Food	cake, fish, lunch, chicken, dinner, meat, bread, rabbit
Institution	school, state, university, police, company
Person	people, man, woman, john, jenny

Table : Semantic roles and their lexical realizations.

## Reconstruction abilities II

Lexical unit	Semantic frames
eat	(Person, Food), (Animal, Food)
cook	(Person, Food)
chew	(Animal, Food), (Person, Food)
bite	(Animal, Food), (Animal, -)
paint	(Person, Person), (Person, -)
teach	(Person, Person), (Institution, Person), (Person, Animal)
love	(Person, Person)
hire	(Institution, Person)
pay	(Institution, Person)
produce	(Institution, Food)
buy	(Institution, Food), (Person, Food), (Person, Animal)
feed	(Person, Animal), (Person, Person)
train	(Person, Animal)

# Reconstruction abilities III



# Conclusions

- LDA-Frames – unsupervised model for generating probabilistic semantic frames
- Non-parametric LDA-Frames
- Useful for linguists
- Many practical applications: statistical thesaurus, word sense disambiguation, machine translation, genetics, etc.

**home page:** <http://nlp.fi.muni.cz/projekty/lda-frames/>

**code:** <http://code.google.com/p/lda-frames/>

**e-mail:** [jiri.materna@firma.seznam.cz](mailto:jiri.materna@firma.seznam.cz)

**twitter:** @JiriMaterna

Thank you for your attention.