#### THE VERB ARGUMENT BROWSER

#### Bálint Sass

sass.balint@itk.ppke.hu

Péter Pázmány Catholic University, Budapest, Hungary

11<sup>th</sup> International Conference on Text, Speech and Dialog 8-12 September 2008, Brno





entence model VCs as coll. Usage & examples Applications Generalization

## **PREVIEW**

A corpus query tool for expressions like ...

- verb subcategorization frames
- institutionalized phrases
- light verb constructions
- idiomatic verbal expressions
- figures of speech
- $\rightarrow$  common property: verb + arguments  $\rightarrow$  uniform framework

Motivation: to help in manually building lexical resources

Future work: apply the methodology to other languages





- SENTENCE MODEL
- 2 VERBAL CONSTRUCTIONS AS COLLOCATIONS
- **3** Usage & examples
- 4 APPLICATIONS
- **5** GENERALIZATION





- SENTENCE MODEL
- VERBAL CONSTRUCTIONS AS COLLOCATIONS
- 3 USAGE & EXAMPLES
- 4 APPLICATIONS
- **5** GENERALIZATION



• Basic unit: simple sentence or clause.

A lány váll-at von. the girl shoulder-ACC pull. 'The girl shrugs her shoulder.'

Clause = verb + set of arguments

verb=von NOM=lány ACC=váll verb=shrug SUBJ=girl OBJ=shoulder

- Positions: defined . . .
  - syntactically: order (in English)
  - morphologically: case markers (in Hungarian)





• Basic unit: simple sentence or clause.

A lány váll-at von. the girl shoulder-ACC pull. 'The girl shrugs her shoulder.'

Clause = verb + set of arguments

verb=von NOM=lány ACC=váll verb=shrug SUBJ=girl OBJ=shoulder

- Positions: defined . . .
  - syntactically: order (in English)
  - morphologically: case markers (in Hungarian)





• in Hungarian: 20 different case markers in English: usually prepositions

case marker	case	abbr.	English
-Ø	nominative	MOM	word order
-t	accusative	ACC	word order
-bAn	inessive	INE	<i>in</i> -phrase
-rÓl	delative	DEL	<i>from</i> -phrase <sup>1</sup>
-bÓl	elative	ELA	from-phrase <sup>2</sup>
			•







Sentence model VCs as coll. Usage & examples Applications Generalization

### **EXAMPLES**

Az emberek az időjárás-ról beszélnek. the people the weather-DEL talk. 'People talk about the weather.'

verb=beszél NOM=ember DEL=időjárás verb=talk SUBJ=people ABOUT=weather

Péter fél az ismeretlen-től. Peter fear the unknown-ABL. 'Peter fears of the unknown.'

verb=fél NOM=Péter ABL=ismeretlen verb=fear SUBJ=Peter OF=unknown



Sentence model VCs as coll. Usage & examples Applications Generalization

## **EXAMPLES**

Az emberek az időjárás-ról beszélnek. the people the weather-DEL talk. 'People talk about the weather.'

verb=beszél NOM=ember DEL=időjárás verb=talk SUBJ=people ABOUT=weather

Péter fél az ismeretlen-től. Peter fear the unknown-ABL. 'Peter fears of the unknown.'

verb=fél NOM=Péter ABL=ismeretlen verb=fear SUBJ=Peter OF=unknown



Sentence model VCs as coll. Usage & examples Applications Generalization

# **EXAMPLES**

Az emberek az időjárás-ról beszélnek. the people the weather-DEL talk. 'People talk about the weather.'

verb=beszél NOM=ember DEL=időjárás verb=talk SUBJ=people ABOUT=weather

Péter fél az ismeretlen-től. Peter fear the unknown-ABL. 'Peter fears of the unknown.'

verb=fél NOM=Péter ABL=ismeretlen verb=fear SUBJ=Peter OF=unknown



### FIXED AND FREE POSITIONS

Hogy jöttek lét-re az első csillagok? how came existence-SUB the first stars? 'How the first stars came into existence?'

*verb=*jön *SUB=*lét *NOM=*csillagok *verb=*come *INTO=*existence *SUBJ=*stars

- fixed position:
   cannot change the word without changing the meaning
- free position:
   can change the word without changing the meaning





### FIXED AND FREE POSITIONS

Hogy jöttek lét-re az első csillagok? how came existence-SUB the first stars? 'How the first stars came into existence?'

*verb*=jön *SUB*=lét *NOM*=csillagok *verb*=come *INTO*=existence *SUBJ*=stars

- fixed position:
   cannot change the word without changing the meaning
- free position:
   can change the word without changing the meaning





### MULTI WORD VERBS

lét-re jön existence-SUB come 'come into existence'

- multi word verb: verb stem + fixed position(s)
- separate meaning
- own argument structure

```
rész-t vesz bAn
part-ACC take INE
'take part in sq'
```



sentence = verb + set of arguments representation of arguments: position + lemma

i.e.

*verb*=jön *SUB*=lét *NOM*=csillagok *verb*=come *INTO*=existence *SUBJ*=stars





# CORPUS PREPARATION

*Input:* Hungarian National Corpus (POS-tagged and disambiguated)

- clause detection
  - regexps based on conjunction and punctuation patterns
- verb normalization
  - e.g. separated verbal prefixes attached
- noun phrase chunking
  - → case and lemma of the head of argument phrases
- → representation according to the model





- 1 SENTENCE MODEL
- 2 VERBAL CONSTRUCTIONS AS COLLOCATIONS
- 3 USAGE & EXAMPLES
- 4 APPLICATIONS
- **5** GENERALIZATION



We search for collocations in the space of these structures:

verb=jön SUB=lét NOM=csillagok
verb=come INTO=existence SUBJ=stars

#### **IDEA**

Apply an association measure taking ...

- the lemma in one particular position as one unit,
- all other parts of the verb frame as the other unit of the collocation.





We search for collocations in the space of these structures:

verb=jön SUB=lét NOM=? verb=come INTO=existence SUBJ=?

#### **IDEA**

Apply an association measure taking ...

- the lemma in one particular position as one unit,
- all other parts of the verb frame as the other unit of the collocation.





The Verb Argument Browser can answer the following typical research question:

- What are the salient words which can appear in a free position of a given verb frame?
- What are the most important collocates of a given verb (or verb frame) in a particular morphosyntactic position?

Association measure: salience (adjusted mutual information)

$$S(x,y) = log_2 f(y) \cdot log_2 N \frac{f(x,y)}{f(x) \cdot f(y)}$$





Important property of the Verb Argument Browser:

It can treat not just a single word but a whole verb frame (a verb together with some arguments) as one unit in collocation extraction.

#### It can collect ...

- salient subjects of a verb,
- salient objects of a given verb—subject pair,
- salient locatives of a given verb—subject—object triplet . . .





- 1 SENTENCE MODEL
- VERBAL CONSTRUCTIONS AS COLLOCATIONS
- 3 USAGE & EXAMPLES
- 4 APPLICATIONS
- **5** GENERALIZATION



- Hungarian National Corpus integrated (187 million running words)
- response times: a few seconds

Corpus: Hungarian National Corpus  Verb:  ker  No:   Case/postposition:  ACC   No:   Argument lemma:   No:   Case/postposition:  ABL   No:   Argument lemma:   No:   String:   Full sentence coverage.	Distribution:
Full sentence coverage:	





- Hungarian National Corpus integrated (187 million running words)
- response times: a few seconds

Corpus:   Hungarian National Corpus	Distribution:
Full sentence coverage:	





- Hungarian National Corpus integrated (187 million running words)
- response times: a few seconds

Vertic   ker   No:   Case/postposition:   ACC   No:   Argument lemma:   Cose/postposition:   ABL   Argument lemma:   ABL		No: Case/postposition: ACC No: No: Case/postposition: ABL No: No: String:	-	Distribution:
---	--	---	---	---------------





- Hungarian National Corpus integrated (187 million running words)
- response times: a few seconds

	Corpus:   Hungarian National Corpus   Verb:   ker   No:   Case/postposition   ACC   No:   Argument lemma:   No:   Case/postposition   ABL   No:   Argument lemma:   No:   String:   Verbis sentence coverage:	Distribution:
--	---	---------------



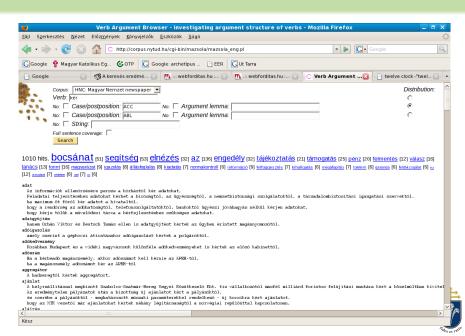


- Hungarian National Corpus integrated (187 million running words)
- response times: a few seconds

Corpus:   Hungarian National Corpus	Distribution:
Search	







kér t tól ask ACC ABL 'ask sy sg'

verb=kér ABL=? ACC=?
verb=ask INDIR=? OBJ=?



kér t tól ask ACC ABL 'ask sy sg'

```
verb=kér ABL=? ACC=?
verb=ask INDIR=? OBJ=?
```

### Result: (Most salient direct objects:)

- bocsánat 'forgiveness'
- segítség 'help'
- elnézés also 'forgiveness'
- engedély 'permission'
- . . . .





kér tól ask ACC ABL 'ask sy sg'

```
verb=kér
       ABL=?
                ACC=?
verb=ask INDIR=?
                OBJ=?
```

### Result: (Most salient direct objects:)

- bocsánat 'forgiveness'
- segítség 'help'
- elnézés also 'forgiveness'
- engedély 'permission'

# for English?

- question
- favour

...





```
vesz figyelem-bA t
take consideration-ILL ACC
'take sg into consideration'
```

```
verb=veszILL=figyelemACC=?verb=takeINTO=considerationOBJ=?
```



vesz figyelem-bA t take consideration-ILL ACC 'take sg into consideration'

verb=vesz ILL=figyelem ACC=?
verb=take INTO=consideration OBJ=?

### Result: (Most salient direct objects:)

- szempont 'aspect'
- érdek 'interest'
- vélemény 'opinion'
- ...

for English?

Probably the same.



ad t give ACC 'give sg'

```
verb=ad ACC=?
verb=give OBJ=?
```





```
ad t
give ACC
'give sg'
```

```
verb=ad ACC=?
verb=give OBJ=?
```

### Result: (Most salient direct objects:)

- hang 'voice' → 'to give voice to sg'
- hír 'news' → to give news ~ 'to report'
- igaz 'true'  $\rightarrow$  to give true  $\sim$  'to take sy's side'
- ...
- → multi word verbs





üt ∅ strike NOM 'sg strikes'

verb=üt NOM=? verb=strike SUBJ=?

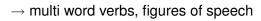


üt ∅ strike NOM 'sg strikes'

verb=üt NOM=? verb=strike SUBJ=?

### Result: (Some salient subjects:)

- óra 'clock' → 'The clock strikes twelve.'
- forint → 10 Ft strikes his palm. ~ 'He receives 10 Ft.'
- $k \ddot{o}$  'stone'  $\rightarrow$  Üsse  $k \ddot{o}!$  Let a stone strike it!  $\sim$  'It does not matter.'
- ...





# COLLECTING MWVS

Important property of the Verb Argument Browser:

Investigating a specific position, the tool provides constructions with this position *fixed* if there is any such construction (e.g. light verb constructions, idiomatic verbal expressions, figures of speech).

- 'kick' + OBJ → 'bucket'
- 'eat' + OBJ → some kinds of food

Verbal expressions with fixed position(s) are frequent, they are not to be ignored, they should be included in language models.



- SENTENCE MODEI
- VERBAL CONSTRUCTIONS AS COLLOCATIONS
- 3 USAGE & EXAMPLES
- 4 APPLICATIONS
- **5** GENERALIZATION



### **APPLICATIONS**

- lexical database development of a Hungarian to English machine translation system
  - → http://www.webforditas.hu
- searching for MWVs to include them into the Hungarian WordNet

- lexicography
- language teaching





entence model VCs as coll. Usage & examples Applications Generalization

# FUTURE WORK

We are planning to create a *Hungarian verb frame frequency dictionary* based on this tool.

If you specify a verb frame, the Verb Argument Browser tells which are the important lemmas in a chosen position.

#### **QUESTION**

How to collect automatically all important constructions of a verb?





- SENTENCE MODEL
- VERBAL CONSTRUCTIONS AS COLLOCATIONS
- 3 USAGE & EXAMPLES
- 4 APPLICATIONS
- **5** GENERALIZATION



entence model VCs as coll. Usage & examples Applications Generalization

#### GENERALIZATION

The database can be anything which fits the model:
 a bigger unit which has positions and
 these positions can be filled by particular items.

 It is possible to use the methodology to investigate argument structure of adjectives or nouns.

The sentence model is in essence language independent.
 The methodology can be extended to other languages, if a shallow parsed, adequately processed corpus is available.





#### SUMMARY

- Verb Argument Browser
  - sentence model + collocation extraction
  - → important verbal constructions
- language independent methodology
  - available for Hungarian:

http://corpus.nytud.hu/vab

(username: tsd; password: vab)

• ...other languages?
Contact: sass.balint@itk.ppke.hu





#### SUMMARY

- Verb Argument Browser
  - sentence model + collocation extraction
  - → important verbal constructions
- language independent methodology
  - available for Hungarian:

http://corpus.nytud.hu/vab

(username: tsd; password: vab)

• ...other languages?
Contact: sass.balint@itk.ppke.hu

Thank you for your attention!

