

# 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



# PREVIEW

A corpus query tool for expressions like ...

- verb subcategorization frames
- institutionalized phrases
- light verb constructions
- idiomatic verbal expressions
- figures of speech

→ *common property*: verb + arguments → *uniform framework*

*Motivation*: to help in manually building lexical resources

*Future work*: apply the methodology to other languages



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



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



# SENTENCE MODEL

- *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)



# SENTENCE MODEL

- *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)



# SENTENCE MODEL

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

case marker	case	abbr.	English
-∅	nominative	NOM	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	<i>from</i> -phrase <sup>2</sup>
...			

# 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



# 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

# 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 sg'



# SENTENCE MODEL

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

i.e.

<i>verb=jön</i>	<i>SUB=lét</i>	<i>NOM=csillagok</i>
<i>verb=come</i>	<i>INTO=existence</i>	<i>SUBJ=stars</i>



# 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





# VERBAL CONSTRUCTIONS AS COLLOCATIONS

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.

# VERBAL CONSTRUCTIONS AS COLLOCATIONS

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.



# VERBAL CONSTRUCTIONS AS COLLOCATIONS

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)

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



# VERBAL CONSTRUCTIONS AS COLLOCATIONS

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



# USAGE

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



Corpus:

Verb:

No:  Case/postposition:  No:  Argument lemma:

No:  Case/postposition:  No:  Argument lemma:

No:  String:

Full sentence coverage:

Distribution:



# USAGE

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



Corpus:

Verb:

No:  Case/postposition:  No:  Argument lemma:

No:  Case/postposition:  No:  Argument lemma:

No:  String:

Full sentence coverage:



Distribution:



# USAGE

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



Corpus:

Ver:

No:  Case/postposition:  No:  Argument lemma:

No:  Case/postposition:  No:  Argument lemma:

No:  String:

Full sentence coverage:

Distribution:





# USAGE

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



Corpus:

Verb:

No:  Case/postposition:  No:  Argument lemma:

No:  Case/postposition:  No:  Argument lemma:

No:  String:

Full sentence coverage:

Distribution:



# USAGE

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



Corpus: Hungarian National Corpus

Verb: kér

No:  Case/postposition: ACC

No:  Argument lemma:

No:  Case/postposition: ABL

No:  Argument lemma:

No:  String:

Full sentence coverage:

Search

Distribution:



Verb Argument Browser - investigating argument structure of verbs - Mozilla Firefox

Fájl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Súgó

http://corpus.nyud.hu/cgi-bin/mazsola/mazsola\_eng.pl

Google Magyar Katolikus Eg... OTP Google: archetipus ... EER Ut Tarra

Google A keresés eredm... webforditas.hu ... webforditas.hu ... Verb Argument ... twelve clock -"twel...

Corpus: HNC: Magyar Nemzet newspaper

Verb: kér

No:  Case/postposition: ACC No:  Argument lemma:

No:  Case/postposition: ABL No:  Argument lemma:

No:  String:

Full sentence coverage:

Distribution:

1010 hits. [bocsánat](#) [51] [segítség](#) [53] [elnézés](#) [32] [az](#) [136] [engedély](#) [32] [tájékoztatás](#) [21] [támogatás](#) [25] [pénz](#) [20] [felmentés](#) [12] [válasz](#) [16] [tanács](#) [13] [forint](#) [16] [magyarulat](#) [9] [igazolás](#) [8] [állásfoglalás](#) [8] [kiadás](#) [7] [normakontroll](#) [6] [információ](#) [9] [figyelgetés](#) [7] [kihallgatás](#) [6] [menyőpótlás](#) [7] [tölem](#) [6] [garancia](#) [6] [felkutatás](#) [6] [az](#) [12] [szavast](#) [7] [szokas](#) [6] [am](#) [7] [=](#) [6]

adatt

Az információk ellenőrzésére perze a körháztól kér adatokat, Feladatai teljesítéséhez adatokat kérhet a bíróságtól, az ügyészségtől, a nemzetbiztonsági szolgálatoktól, a társadalombiztosítási igazgatási szervektől, ha maximum öt főről kér adatot a hivataltól.

hogy a rendőrség az adathozságtól, telefonszolgálatoktól, bankoktól ügyészi jóváhagyás nélkül kérjen adatokat, hogy kérje tőlük: a művelődési tárca a bérfejlesztéshez szükséges adatokat.

adattgyűjtés

hanes Orbán Viktor és Deutch Tamás ellen is adattgyűjtést kértek az ügyben érintett magánnyomozótól.

adóigazolás

anely szerint a gépkocsi átíratásakor adóigazolást kértek a polgároktól.

adókedvezmény

Korábban Budapest és a vidéki nagyvárosok különféle adókedvezményeket is kértek az előző kabinettől,

adószán

Ha a bérbeadó magánszemély, akkor adószámat kell kérnie az APER-től, ha a magánszemély adószámat kér az APER-től

aggregátor

A hadneregtről kértek aggregátort,

ajánlat

A helyreállításával megbízott Szabolcs-Szatmár-Bereg Megyei Közműkezelő Hht. tíz vállalkozótól másfél milliárd forintot felújítási munkára kért a közelműltban kivitel. Az eredménytelen pályázatok után a bizottság új ajánlatot kért a pályázóktól.

és cserébe a pályázóktól - meghatározott műszaki paraméterekkel rendelkező - új kocsikra kért ajánlatot.

hogy az MTK vezetői már ajánlatokat kértek néhány légitársaságtól a norvégiai repülőtérről kapcsolatosan,

Aláírás

Kész



*Query:*

kér t tól  
 ask ACC ABL  
 'ask sy sg'

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

*Query:*

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'
- ...

*Query:*

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'
- ...

*for English?*

- question
- favour
- ...



Query:

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

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



Corpus: HNC: Magyar Nemzet newspaper ▾  
 Verb: vesz  
 No:  Case/postposition: ILL No:  Argument lemma: figyelem  
 No:  Case/postposition: ACC No:  Argument lemma:  
 No:  String:  
 Full sentence coverage:   
 Search

Distribution:



*Query:*

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.





Query:

ad t  
give ACC  
'give sg'

*verb=ad* ACC=?

*verb=give* OBJ=?



Query:

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' → to give true ~ 'to take sy's side'
- ...

→ multi word verbs

*Query:*

üt      ∅  
strike    NOM  
'sg strikes'

*verb=üt      NOM=?*  
*verb=strike    SUBJ=?*

Query:

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.'
- kő – 'stone' → Üsse kő! – Let a stone strike it! ~ 'It does not matter.'
- ...

→ multi word verbs, figures of speech

# 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.



- 1 SENTENCE MODEL
- 2 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



## 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?





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

