# Finding English Equivalents of Hungarian Light Verb Constructions

**István Nagy T.** [1] **, Veronika Vincze**[2]
[1]Collokia LLC
e-mail: istvan@collokia.com
[2]MTA-SZTE, Hungarian Academy of Sciences
Research Group on Artificial Intelligence,
e-mail: vinczev@inf.u-szeged.hu

WG3

## Introduction

•The meaning of light verb constructions (LVCs) can only partially be computed on the basis of the meanings of their parts and the way they are related to each other (semi-compositionality)
•The result of translating their parts literally can hardly be considered as the proper translation of the original expression:

***döntést hoz decision-ACC bring – *bring a decision***
****döntést tesz decision-ACC make – make a decision***

**Goal**: train a machine learning-based model to automatically identify the English equivalents of Hungarian LVCs based on a Hungarian – English parallel corpus.
Hungarian:
*Látták, hogy bemászik az ablakon, úgyhogy nem* see-PAST-3PL, that in.climb-3SG the window-SUP, so not *lehetett titokban tartani.*
can-PAST-3SG secret-INE keep-INF.
English: *She was seen climbing through the window, so it couldn't be kept a secret.*

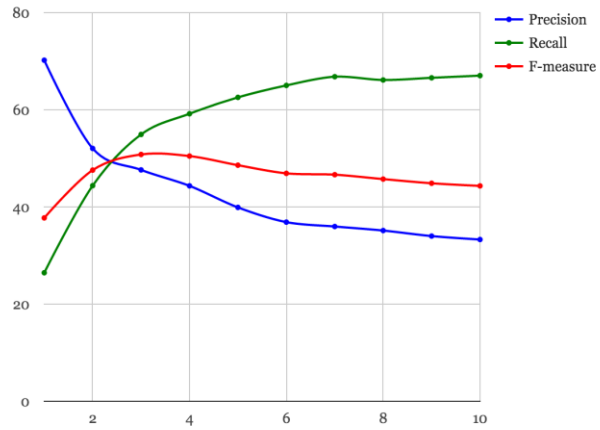
***titokban tartani != climbing through the window***
***titokban tartani = kept a secret***

## Corpus

•Apply the SzegedParalellFX [1] parallel corpus where LVCs were manually annotated in Hungarian and English as well.
•We made use of the gold standard annotation of Hungarian LVCs, and generated their potential English equivalents.
•From the candidates, a linguist selected the correct translation for the original LVC.

- 1377 LVC on the Hungarian part of the corpus
- 4635 generated potential English equivalents
- 446 English LVC were correct equivalents

| Method | Baseline | Decision Tree |
|---|---|---|
| Precision | 73.68 | 47.63 |
| Recall | 15.69 | 54.93 |
| F-measure | 25.88 | 50.81 |




## Machine Learning-based model

- Investigate how potential translation pairs can be automatically identified
- baseline method: context-free dictionary lookup
  - pairs are correct translations when the two nominal components proved to be translational equivalents in a Hungarian-English dictionary.
- Machine learning-based method with feature set optimized for English–Hungarian LVC detection (morphological, syntactic and lexical features) [2,3].
  - New features: nominal and verbal translation pairs
  - J48 classifier of the WEKA package
- Evaluate the model in 10-fold cross validation manner on the SzegedParallelFX corpus.
- The negative examples were overrepresented therefore we gave extra weight to the positive examples during the training process.
- The efficiency of the machine learning-based method changes with different weights.

## Results

- problematic for the system to identify LVCs which contain verbs rarely used in LVCs:
  ***(nehéz) életet élnek - lead (difficult) lives***
  (difficult) life-ACC live-3PL
- One of the most difficult tasks was when both the English and Hungarian translation units contained LVCs but they but did not correspond to each other:

Hungarian:
*Háromévi várakozás után William Prichard kapitány, az Antilop* of.three.years waiting after William Prichard captain, the Antelope *gazdája, ki a déli vizekre volt indulóban,* owner-3SGPOSS, who the Southern water-PL-SUB was leave-INE, *előnyös ajánlatot tett nekem, és én elfogadtam.* advantageous offer-ACC make-PAST-3SG I-DAT, and I accept-PAST-1SG-OBJ.
English:
*After three years expectation that things would mend, I accepted an advantageous offer from Captain William Prichard, master of the Antelope, who was making a voyage to the South Sea.*

- The quality of the automatic dictionary can be modified by weights depending of the end application. (to create an accurate dictionary vs. get many possible translation pairs)
- Future work: a dictionary of Hungarian-English LVCs can be automatically created

## References

1. Veronika Vincze. 2012. Light Verb Constructions in the SzegedParalellFX English–Hungarian Parallel Corpus. In Proceedings of LREC-2012, pages 2381–2388, Istanbul. ELRA.
2. Veronika Vincze, István Nagy T., János Zsibrita. 2013. Learning to detect English and Hungarian light verb constructions. ACM Transactions on Speech and Language Processing (TSLP) - Special issue on multiword expressions: From theory to practice and use. Part 1, 10(2), Article 6.
3. Veronika Vincze, István Nagy T., and Richárd Farkas. 2013. Identifying English and Hungarian Light Verb Constructions: A Contrastive Approach. In Proceedings of ACL 2013, pages 255–261.