

The annotation of discourse markers in an English-Hungarian parallel corpus: parsing out the complex relationship between formal and functional features

1. Introduction

Motivations:

- there is an extremely complex relationship between the formal and functional properties of DMs (cf. Aijmer 2013: 18);
- the prosodic realization of DMs still doesn't receive enough prominence in DM research (Wichmann et al. 2010: 103);
- there is general agreement in the literature that a contrastive analysis can help tease out the diversity of meaning relations that semantically bleached DMs mark (cf. Mortier & Degand 2009).

Objectives:

- to identify the most relevant formal features of **of course** that help us disambiguate between textual and interpersonal as well as contrastive and non-contrastive functions;
- to use these formal features in order to model the basic discourse functions of **of course** via a decision tree, which can be tested on a larger corpus.

2. Research material: scripted dialogues

- LAC: First four seasons of House (also known as House M. D., © NBC Universal Television) transcripts rather than subtitles
- LBC: Hungarian translations (2 versions) transcripts of dubbed releases
- 837,088 words ~5% (characters' names and stage instructions)
- LAC+LBC1-2 aligned (SDL trados)
- 200 tokens of *of course* (251 tokens per million words ~ 190 tpmws in NOSD)

3. The functional spectrum of *of course*

Previous research on *of course*:

Lewis (2006): (1) emphatic yes, (2) concession, (3) background, (4) end of list
 Aijmer&Simon-Vandenberg (2004): (1) evidential, (2) interpersonal, (3) indeterminate
 Furkó (2007): (1) conversation management, (2) information management, (3) story structure, (4) interpersonal uses

Basic functions to be disambiguated in the corpus	Sample utterances
INTERP (=interpersonal and interactional function is more salient than textual function) 60% of tokens	Amy: I'm just a little nauseous I umm... I think I ate too much. Can we take a break? Henry: Of course.
CONTR (=contrastive and concessive textual function including reformulation) 30% if tokens	Wilson: I'm not gonna date a patient's daughter. House: Very ethical. Of course , most married men would say they don't date at all.
ADD (=additive/non-contrastive textual function including result, elaboration and justification) 10% of tokens	Let's see, your stomach has the deep-seated feelings of abandonment written all over it which points towards sexual abuse. Well a fear of hospitals; that points to a more specific traumatic event, so I'm going to say-- your mom, in the hospital with a candlestick. And by candlestick, of course , I mean inherited OTC deficiency.
Marginal resultative (interactional function is also salient: SPP+int. pros.+mid-turn)	House: Are you tired? Dan: Sometimes. Dad: He never sleeps! Of course he's tired.

Advantages and disadvantages of using scripted/dramatised conversations

PROS

- Wide range of functions, similar co-occurrence patterns (cf. Furkó: 2011)
- Data is more neat (fewer fillers, overlaps, lexical search, false start, etc. functions)
- Easier to transcribe
- Readily available translations
- Translators aim for maximum explicitness (cf. House:2015)

CONS

- Different frequencies of particular functions
- Data is spoken-like, rather than naturally-occurring discourse (less prototypical but authentic, cf. Chovanec: 2011, Dynel: 2011)

4. Annotation tags

Formal properties:

- Prosody (stress and integration)
 - Relevant co-text (DM clusters, collocations)
- Speaker roles (doctor to doctor, doctor to patient, exchange bw friends, etc.)
 - Speech act of the host utterance / preceding utterance
 - Position in the utterance (initial, medial, final)
 - Position in the turn (initial, medial, final)
- The host unit's position in conversational structure (second pair part/reaction/loose connection)

Functional properties:

- Textual functions (contrast, concession, addition, result, elaboration, justification)
 - Stance (irony, boost, uncertainty)
- The annotator's confidence in tagging the textual function of the DM
 - The annotator's confidence in tagging the speaker's stance

References

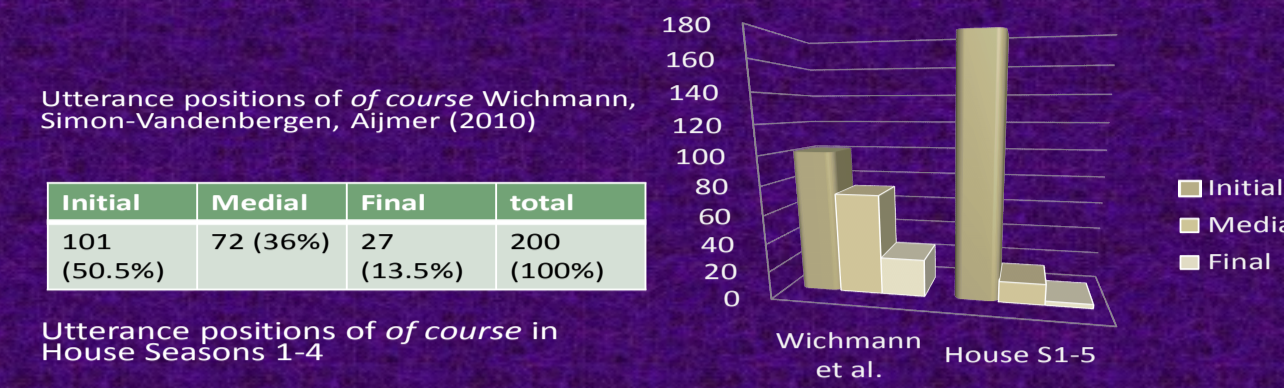
- Aijmer, K. 2013. *Understanding Pragmatic Markers-A Variational Pragmatic Approach*. Edinburgh: Edinburgh University Press.
- Aijmer, K. Simon-Vandenberg, A-M. 2004. A model and a methodology for the study of pragmatic markers: the semantic field of expectation. *Journal of Pragmatics* 36: 1781-1805.
- Chovanec, J. 2011. 'Humour in quasi-conversations - Constructing fun in online sports journalism'. In: Dynel, M. (ed.) *The Pragmatics of Humour across Discourse Domains*. Amsterdam/Philadelphia: John Benjamins, 243-264.
- Dynel, M. 2011. 'Stranger than fiction? - A few methodological notes on linguistic research in film discourse'. *Brno Studies in English* 37(1), 41-61.
- Furkó, B. P. 2007. The status of *of course* as a discourse marker. *HUSSE 8 Conference Proceedings*. <http://husse-esse.hu/wp-content/2007/04/furko-peter-of-course-as-a-dm.doc>.
- Furkó, B. P. 2011. A Contrastive Study of English *of course* and Hungarian *persze*. *Proceedings of the HUSSE 10 Conference, Linguistics Volume*. Hungarian Society for the Study of English, Debrecen.
- House, Juliane. 2015. Translation studies and Pragmatics. Paper presented at *IPRA 2015*.
- Lewis, D. 2006. Discourse markers in English: a discourse-pragmatic view. In *Approaches to Discourse Particles*, edited by Kerstin Fischer, 43-59. Amsterdam: Elsevier.
- Mortier, L. and L. Degand. 2009. 'Adversative discourse markers in contrast'. *International Journal of Corpus Linguistics* 14(3): 339-66.
- Wichmann, A., A-M. Simon-Vandenberg, and K. Aijmer. 2010. "How prosody reflects semantic change: A synchronic case study of *of course*." In *Subjectification, Intersubjectification and Grammaticalization*, edited by Kristin Davids, Lieven Vandelanotte and Hubert Cuyckens, 103-155. Berlin and NY: De Gruyter Mouton.

5. Some results of the quantitative analysis

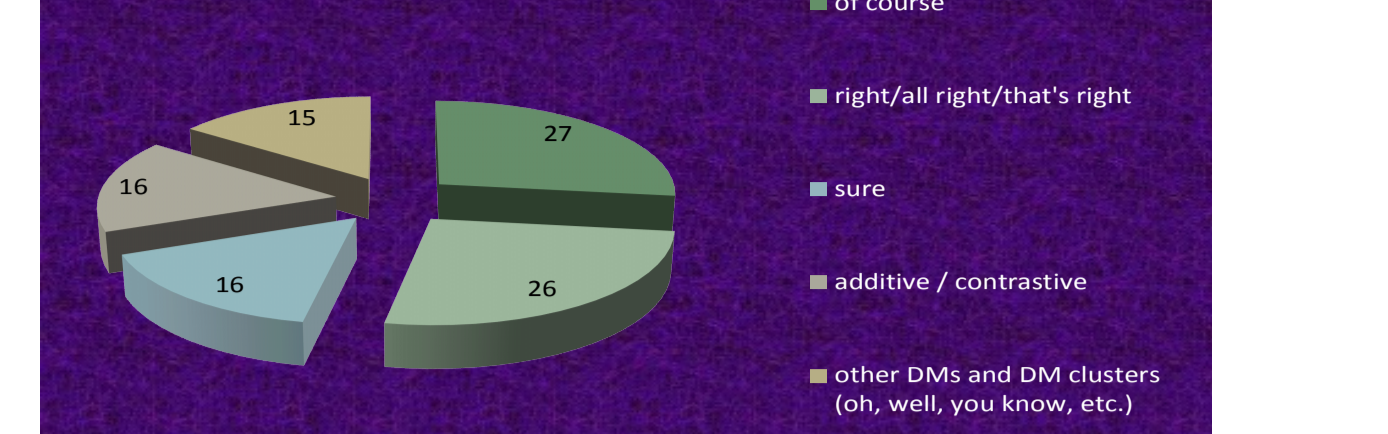
Cluster analysis of 'of course'

- of course I / I'm	40
- of course not	24
- of course it / it's	25
- of course you / you're	17
- of course he	10
- of course the	6
- well of course	6
- No, of course	5
- Yeah, of course	5

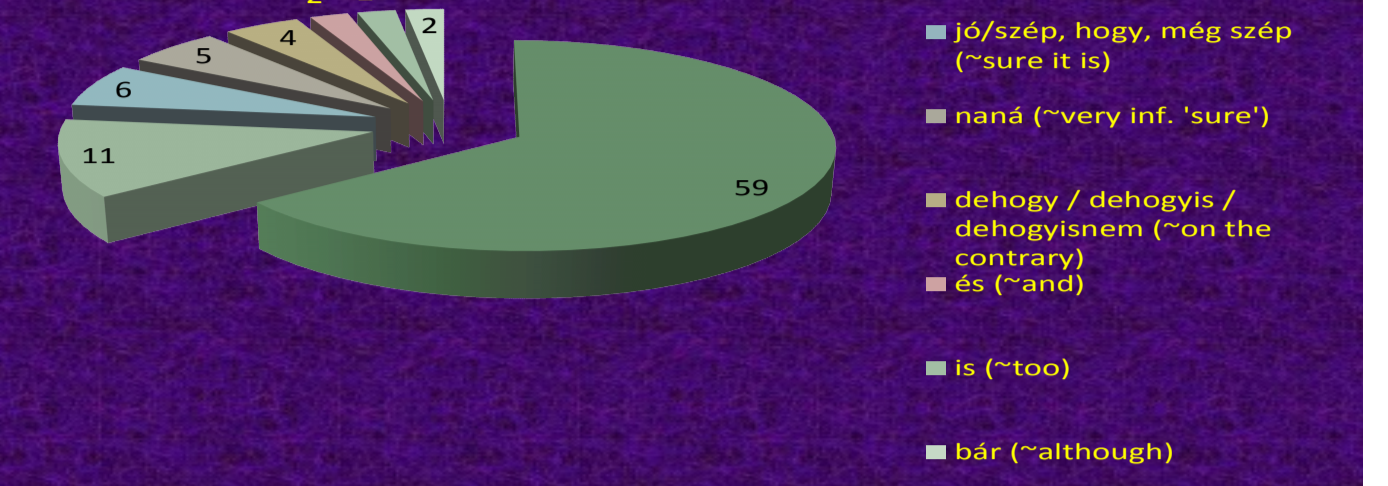
Position of 'of course'



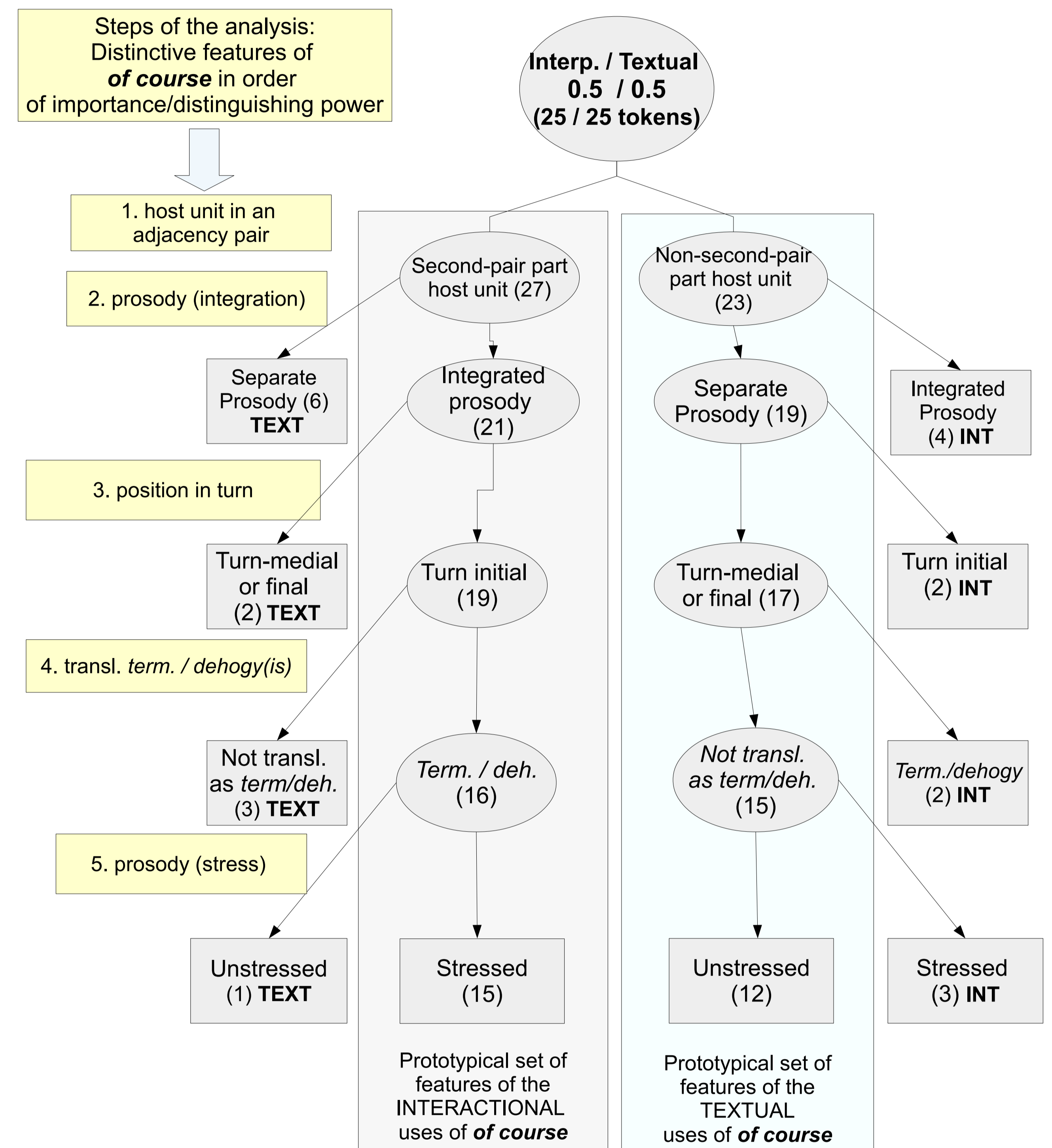
'persze' as a translation equivalent



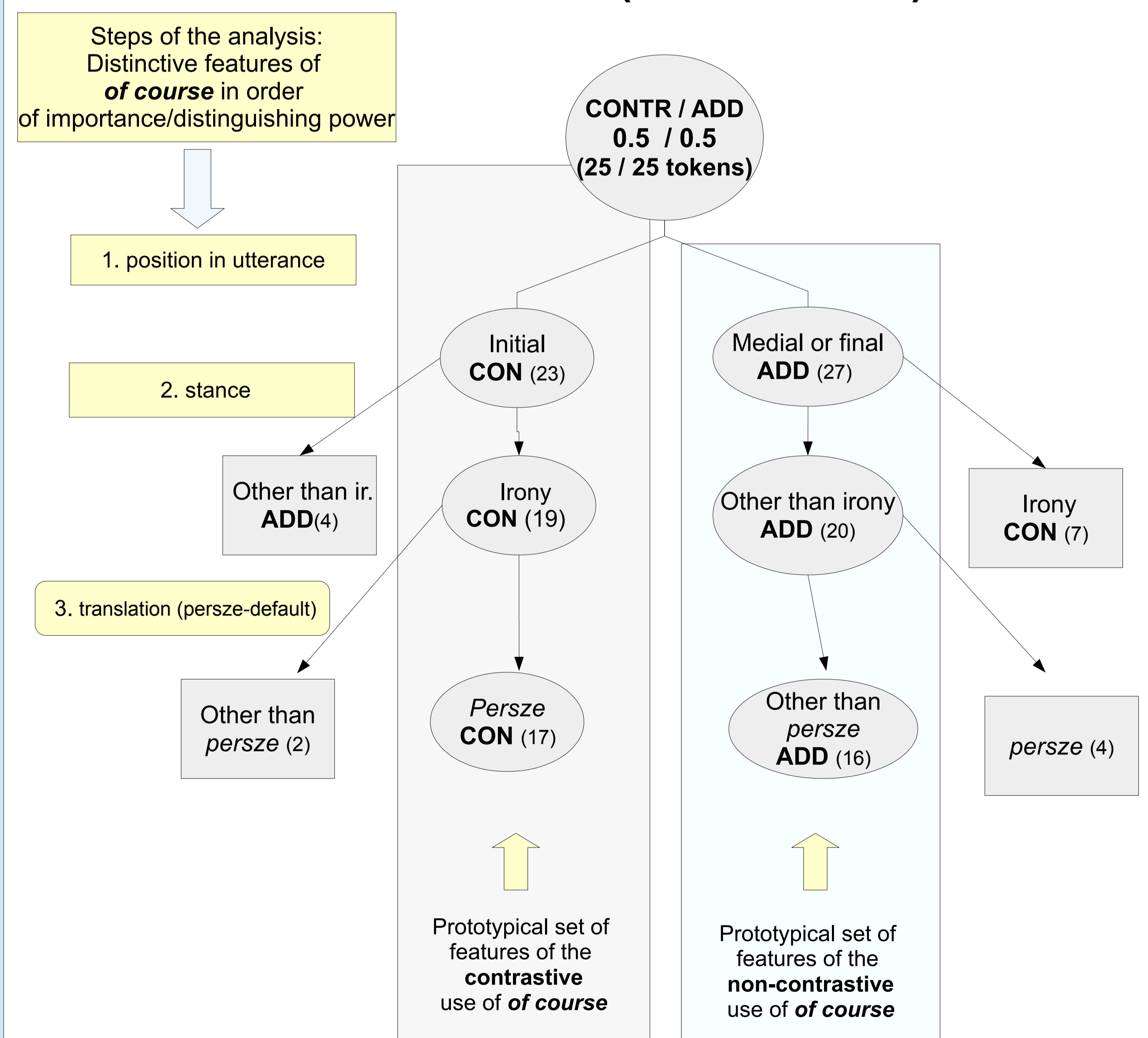
Translations of 'of course'



6. Conclusions 1: A decision tree for the functional disambiguation between interpersonal and discourse uses of *of course*



7. Conclusions 2: A decision tree for the functional disambiguation between contrastive and additive (non-contrastive) uses of *of course*



Péter Bálint Furkó
 Károli Gáspár University of the Reformed Church in Hungary

furko.peter@gmail.com

DiSpOL 2015

Saarland University, Saarbrücken
 1-2 October, 2015