
Cross-linguistic variation: DRDs identification and annotation

Liesbeth Degand
COST Training School
January 2016

From multilingualism to cross-linguistic studies

- ▶ TextLink teams involved with **over 20 languages**
- ▶ Resources and/or linguistic analysis
 - ▶ Written
 - ▶ Czech, Dutch, English, Finnish, French, German, Greek, Italian, Lithuanian, Norwegian, Polish, (Brazilian) Portuguese, Romanian, Turkish,
 - ▶ Spoken
 - ▶ Catalan, English, French, German, Hebrew, Hungarian, Slovene, Spanish
 - ▶ Sign Language
 - ▶ French Belgian Sign Language
- ▶ Multilingual data, not necessarily cross-linguistic!

From multilingualism to cross-linguistic studies

- ▶ Cross-linguistic studies
 - ▶ Written
 - ▶ French/English; German/English; Norwegian/English, Finnish/French...
 - ▶ Spoken
 - ▶ French/English
- ▶ **Challenge:** Perform cross-linguistic analyses on the basis of monolingually developed data

Language comparison

- ▶ Investigating a language implies determining the specific properties/characteristics of that language.
 - ▶ Number and type of DRDs
 - ▶ Onomasiological and semasiological approaches
- ▶ Determine the pattern and limits of variation in human language
 - ▶ Variation in form and function of DRDs
 - ▶ “important variations exist in the number of connectives languages display to express a given relation, even between typologically related languages” (Zufferey & Degand, in press)
- ▶ Locating a language in the space of possibilities
 - ▶ Explicitation and implicitation of discourse relations
 - ▶ Combination of DRDs
 - ▶ Interaction of DRDs with syntax, prosody, thematic structure, ...

Challenges for cross-linguistic analysis

- ▶ Establishing comparability
 - ▶ Categories compared must be of the same type; cf. *tertium comparationis* (TC)
- ▶ Assuming
 - ▶ TC= common platform, against which differences and similarities can be stated
 - ▶ DF=distinctive feature that assumes values DF_1 and DF_2 for L_1 and L_2 respectively
- ▶ Then
 - ▶ L_1 is like L_2 with respect to TC
 - ▶ e.g. [English] _{L_1} is like [French] _{L_2} with respect to [the presence of 2sg personal pronouns]_{TC}.
 - ▶ TC in L_1 is marked by DF_1 , whereas in L_2 TC is marked by DF_2
 - ▶ e.g. [The presence of 2sg personal pronouns]_{TC} in English is marked by [you] _{DF_1} , whereas in French it is marked by [tu, vous] _{DF_2}

Challenges for cross-linguistic DRD analysis

- ▶ Establishing comparability
 - ▶ **Tertium Comparationis** for DRDs (discourse connectives, discourse markers)
 - ▶ “linguistic expression whose primary function lies at the discourse level, i.e. relating their host utterance to the discourse situation (...) contributing to the discourse organization (textual coherence), to the speaker/hearer interaction (interpersonal meanings), and/or to speaker attitudes (epistemic meaning)” (Degand 2014, 151)
 - ▶ lexical items encoding a coherence relation between two abstract objects such as events, states or propositions (PDTB Research Group, 2007)
- ▶ Assuming
 - ▶ Operational definition of DRDs (cf. Bolly et al. 2015, forthc.)
 - ▶ Distinctive realisation of DRDs in L₁ and L₂ → **identification of DRDs!**
- ▶ Then
 - ▶ [English]_{L1} is like [French]_{L2} with respect to [the existence of DRDs]_{TC}.
 - ▶ [The existence of DRDs]_{TC} in English is marked by [*so, then, because, I mean, ...*]_{DF1}, whereas in French it is marked by [*tu vois, parce que, alors, donc, ...*]_{DF2}

Challenges for cross-linguistic causal DRD analysis

- ▶ Establishing comparability
 - ▶ **Tertium Comparationis** for **causal** DRDs (discourse connectives, discourse markers)
- ▶ Assuming
 - ▶ TC = "A causal connective (DRD) is a (grammaticalised) lexical item (conjunction or adverbial?) encoding a causal coherence relation between two clauses"
 - ▶ Causal connectives are realised differently in English and French (and ...)
- ▶ Then
 - ▶ [English]_{L1} is like [French]_{L2} with respect to [the existence of causal DRDs]_{TC}.
 - ▶ [The existence of causal DRDs]_{TC} in English is marked by [*so, because, therefore, since...*]_{DF1}, whereas in French it is marked by [*donc, parce que, car, puisque, du coup, ...*]_{DF2}
 - ▶ [The presence of **subjective backward** causal DRDs]_{TC} in English is marked by [*because, since*]_{DF1}, whereas in French it is marked by [*parce que, car, puisque, du coup*]_{DF2}
 - ▶ ...

Challenges for contrastive linguistics

- ▶ Haspelmath, Martin (2010). 'Comparative concepts and descriptive categories', Language 86: 663-687.
- ▶ "Comparability of incommensurable systems"
 - ▶ Typical example: Can the category of 'aspect' in Russian be compared to the category of 'aspect' in English?
 - ▶ TextLink → Can the (comparative) concepts "discourse relation", "causal relation", "expansion" be compared in English, French, Finnish, Turkish, ...
→ yes (we hope so!)
 - ▶ Can the descriptive category DRD be compared in English, French, Finnish, Turkish, ... → partially (I think!)
 - ▶ → **Comparative concepts and descriptive categories are different** kinds of entities

Challenges for Contrastive linguistics

- ▶ Every language should be described in its own terms
- ▶ A comparative concept is something made by a linguist
... but so is the descriptive category
- ▶ A language-particular study is not merely a matter of checking a pre-established list of comparative concepts ... but that does not mean that one should start from 'categorical scratch' for each 'new' language
 - ▶ one can use categories established for other languages – 'pre-established' – 'comparative' concepts, and also 'descriptive categories'

Doing cross-linguistic (DRD) analysis

- ▶ Comparable data
 - ▶ Data similar in genre and quantity of words over more than one language
 - ▶ E.g. newspaper articles, fiction (original!), blogs, CMC, spontaneous conversation, political address, editorials, scientific articles
 - ▶ Authentic language, same genre (and register), different topics, variety of writers/speakers
- ▶ Parallel data
 - ▶ Translated data: original L1 → translated L2 (and vice-versa)
 - ▶ Fiction (original vs. translation, e.g. Harry Potter), sacred texts, administrative texts, subtitles!, instruction manuals, interpreting (?), ..
 - ▶ Translated language (translationese?), same genre and register, same topic, fewer speakers/writers
- ▶ Post-hoc (comparable)
 - ▶ Select comparable data out of a (heterogeneous) set of data

DRD description on the basis of comparable data

- Contributions in Aijmer, K. & Simon-Vandenberg, A-M (eds.) (2006). *Pragmatic Markers in Contrast*. [Studies in Pragmatics 2]. Amsterdam/Boston: Elsevier.
 - Adverbial connectors in second initial position in English and Swedish
 - *Surely* and its functional counterparts in Spanish
 - *Dutch toch* vs. German *doch*
 - English *now* and Norwegian *nå*
 - *Vraiment* and *really* in contrast
 - Adversative discourse markers in comparable data
 - ...

categorical

semasiological

onomasiological

An example: « en fait » in comparable and parallel data

- Uncover **semantic profile** of adversative DRDs (French/Dutch)
 - understudied
- Combined corpus approach
 - Comparable written data
 - Comparable spoken data
 - Parallel written data
 - Mirror analysis through translation
 - Genre and register variation

Corpus Composition

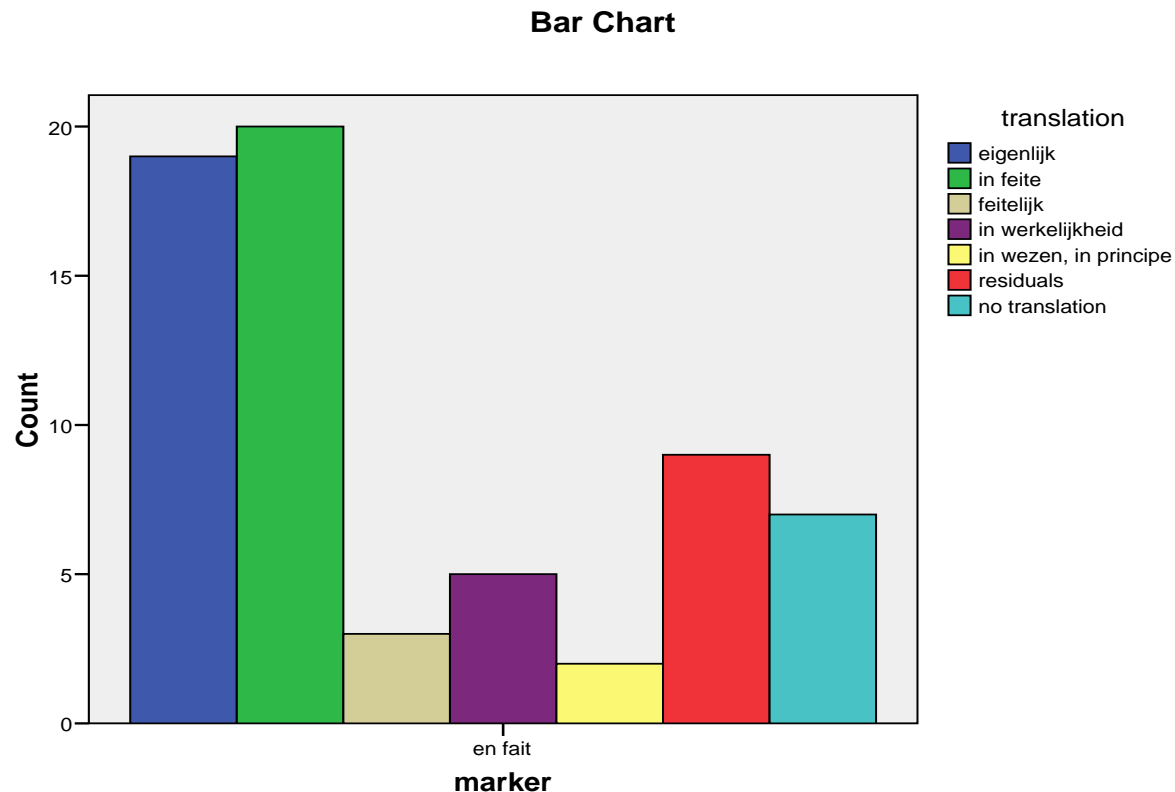
- synchronic **written** comparable data (Fr/Du): 56% newspapers, 30% novels, 14% essays (self-compiled)
- synchronic **spoken** comparable data (Fr/Du): spontaneous face-to-face conversation (Valibel and CGN databases)
- **(diachronic** written comparable data (Fr/Du): literary prose (self-compiled, Frantext))
- synchronic **translation** data (Fr \leftrightarrow Du): newspapers and novels (self-compiled)

		words
translation corpus	Source Dutch	359.810
	Source French	187.763
comparable written corpus	Dutch	1.634.082
	French	1.833.481
comparable spoken corpus	CGN	1.431.545
	Valibel	50.668 (3.483.131)

Table 1. Overview of corpus data used

Clues from translation...

- **HYPOTHESIS:** Fr. *en fait* and Du. *in feite* are translation equivalents



The diagram illustrates the relationships and translation percentages between Dutch words. The central node is **eigenlijk** (striped box). It has several incoming and outgoing arrows with percentages:

- From **in feite** (striped box) via a box containing 30.8% and 42.9%.
- From **in werkelijkheid / wezen / principe, zekere zin** (grey box) via a box containing 15.4%.
- From **en/dans la réalité, vraiment, réellement, à vrai dire** (grey box) via a box containing 9.1%.
- From **En fin d** (grey box) via a box containing 9.1%.
- From **finalement** (striped box) via a box containing 9.1%.
- From **enfin** (striped box) via a box containing 3.2%.

eigenlijk also has outgoing arrows:

- To **in feite** (striped box) via a box containing 42.9%.
- To **in werkelijkheid / wezen / principe, zekere zin** (grey box) via a box containing 10.8%.
- To **en/dans la réalité, vraiment, réellement, à vrai dire** (grey box) via a box containing 47.7%.
- To **au fait, au juste, au fond** (grey box) via a box containing 8%.

There are also direct arrows between **in feite** and **en fait** (striped box) with percentages 30.8% and 42.9%.

A red line highlights a path starting from **in feite**, going down to **en fait**, then right to **eigenlijk**, and finally down to **no translation** (striped box).

- **HOWEVER:** *in feite* is extremely infrequent, especially when compared to *eigenlijk*

	eigenlijk	/ 10.000 words	in feite	/ 10.000 words
translation corpus	83	2,3	9	0,25
written corpus	502	3,1	97	0,6
spoken corpus	3712	25,93	11	0,08

- **SO:** ‘better?’ equivalence between *en fait* and *eigenlijk*

Ensuing Research Questions

- To what extent do *en fait* and *eigenlijk* manifest similar behaviour?
- Are there differences between the spoken and written register, (and through time)?

Meaning description

- Proposal: 4 parameters (tertium comparationis)
 - **Meaning distribution**: weight of pragmatic inferences vs. semantic functions
 - **Nature of host propositions**: fact/action vs. opinion (Pander Maat and Degand 2001)
 - **Scope of *eigenlijk* / *en fait***: local (host proposition q), broad (relation $p \leftrightarrow q$), external (speaker-related)
 - **Position**: intrapropositional, proposition-initial, proposition-final

Meaning of *en fait* (and *eigenlijk*)

- Adversity
 - "contrast between different points of view as these are constructed in language use" (Schwenter 2000, 259)
 - Subclass of **contrast**
 - forces an interpretation in the hearer in which the speaker explicitly asserts (supposed) incompatibility between p and q and indicates his viewpoint as the only relevant (cf. Schwenter 2000, 260)
 - Subclass of **concession** (Malchukov 2004)
 - characteristic of situations in which "someone acknowledges that in highly similar circumstances a mind very similar to one's own draws a valid causal inference, while this inference is actually not valid" (Verhagen 2000, 367).

Meaning distribution

- *en fait* and *eigenlijk* are both very polyfunctional:
 - **explicitation of an implicit opposition** to p
e.g. La fameuse démocratie, rempart supposé de toutes les injustices, reposait **en fait** sur les rapports de force des différents lobbies.
*This famous democracy (...) rested **en fait** (in fact) on the force balance between several lobbies.*
 - **enhancement of an explicit opposition** to p
e.g. Op mijn zwerftochten door de bossen, altijd vergezeld van een rossig hondje, ontmoet ik een jongen die zich Jan noemt **maar eigenlijk** David heet.
*... I met a boy who called himself Jan but **eigenlijk** (actually) was called David.*
 - **counterexpectation**
e.g. Je pense, **en fait**, que les Chinois vont faire l'impossible pour que tout se déroule impeccablement, affirme Bill Clinton.
*I think, **en fait** (in fact/actually), that the Chinese will do the impossible to ...*

- **enhancement** of p, q or part of p/q
 e.g. Vrouwen hebben "aan de wieg van onze beschaving" gestaan. Vrouwen hebben **eigenlijk** alles van betekenis uitgevonden.
*... Women have **eigenlijk** (actually) invented everything of importance.*
- **attenuation** of p, q or part of p/q
 e.g. Lorsque deux corps mous se heurtent et s'immobilisent, leur force est-elle détruite, ou bien s'agit-il seulement d'une apparence, la force s'étant **en fait** dissipée dans les parties menues des corps?
*... the force having **en fait** (in fact) evaporated in the minor parts ...*
- **precision** of p, q or part of p/q
 e.g. Wanneer de last niet dichterbij het lichaam kan gebracht worden, moet je trachten het lichaam, **eigenlijk** het lichaamszwaartepunt, dichterbij de last te brengen.
*... you have to try to bring the body, **eigenlijk** (in fact) the body centre of gravity, closer to the burden*
- **causality** for p (cause, consequence, conclusion)
 e.g. Ces deux formes de récréation sont relativement neuves en Belgique et ont **en fait** été propagées par les Pays-Bas.
*These two forms ... are relatively new in Belgium and have **en fait** (in fact) been introduced through the Netherlands.*
- **metaperspective** (disconnected thought, metacomment, mental leap)
 e.g. Voelen jullie je ook zo zalig?' vraagt Liana en ze drukt precies uit wat wij allemaal voelen. 'Die Jif is **eigenlijk** best een geschikte jongen,' zegt ze eindelijk.
*[expression of feelings] This Jif is **eigenlijk** (actually) a good guy ...*

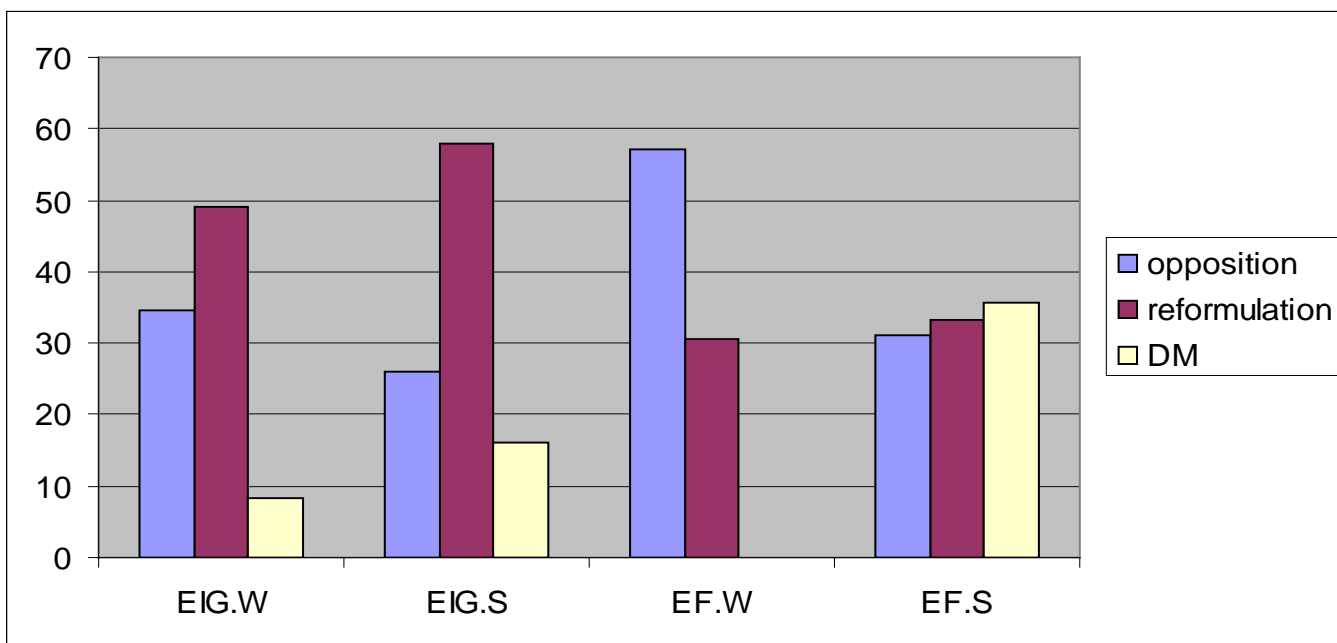
Meanings of opposition

	eigenlijk	en fait
opposition to p	17 (34.7%)	28 (57.1%)
reformulation of p	24 (49%)	15 (30.6%)
metadiscourse marking	4 (8.2%)	absent
no clear association	4 (8.2%)	6 (12.2%)
Total	49	49

Table 6. Meanings expressed between p and q on an ideational level for *eigenlijk* and *en fait*

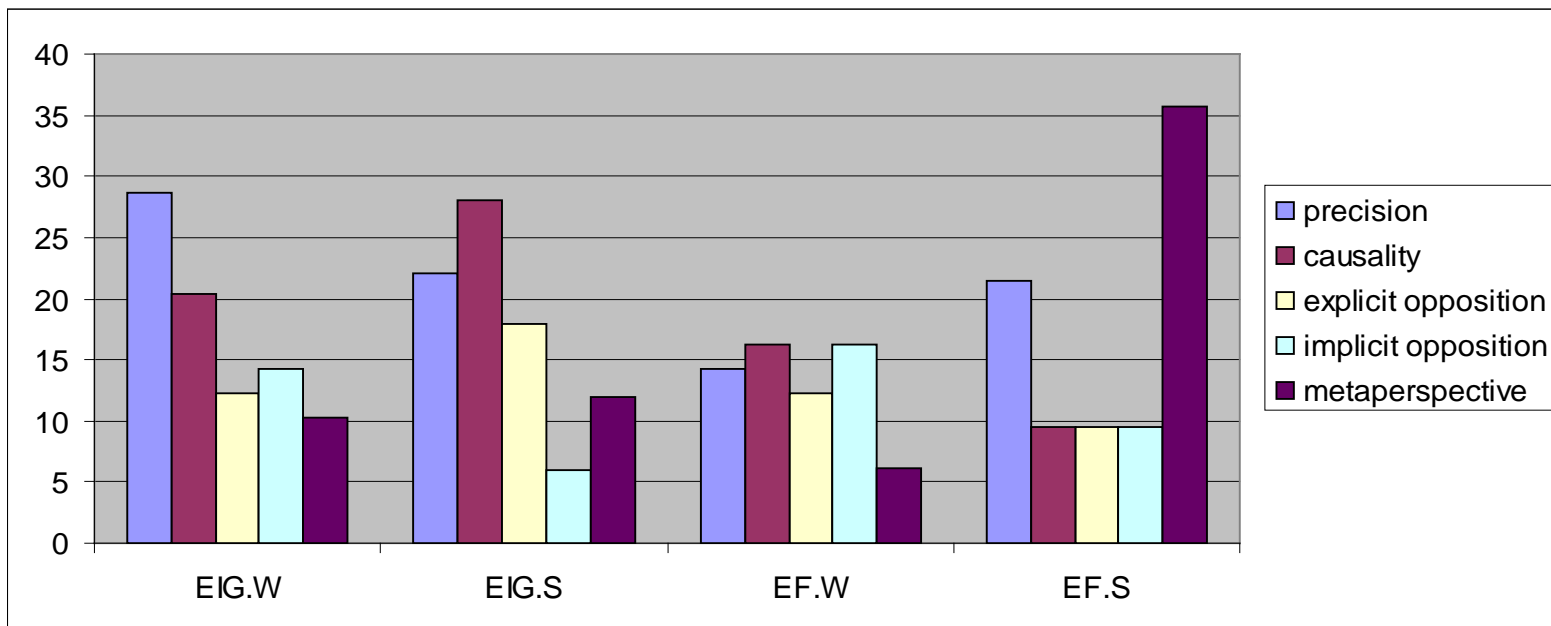
Meaning distribution (1)

- dominant relations:
 - FR vs. DU: *en fait* “oppositive”; *eigenlijk* “reformulative”
 - SP vs. WR:
 - *discourse marking* increases in spoken language, especially in French
 - *reformulation* increases as well, to the detriment of *opposition*

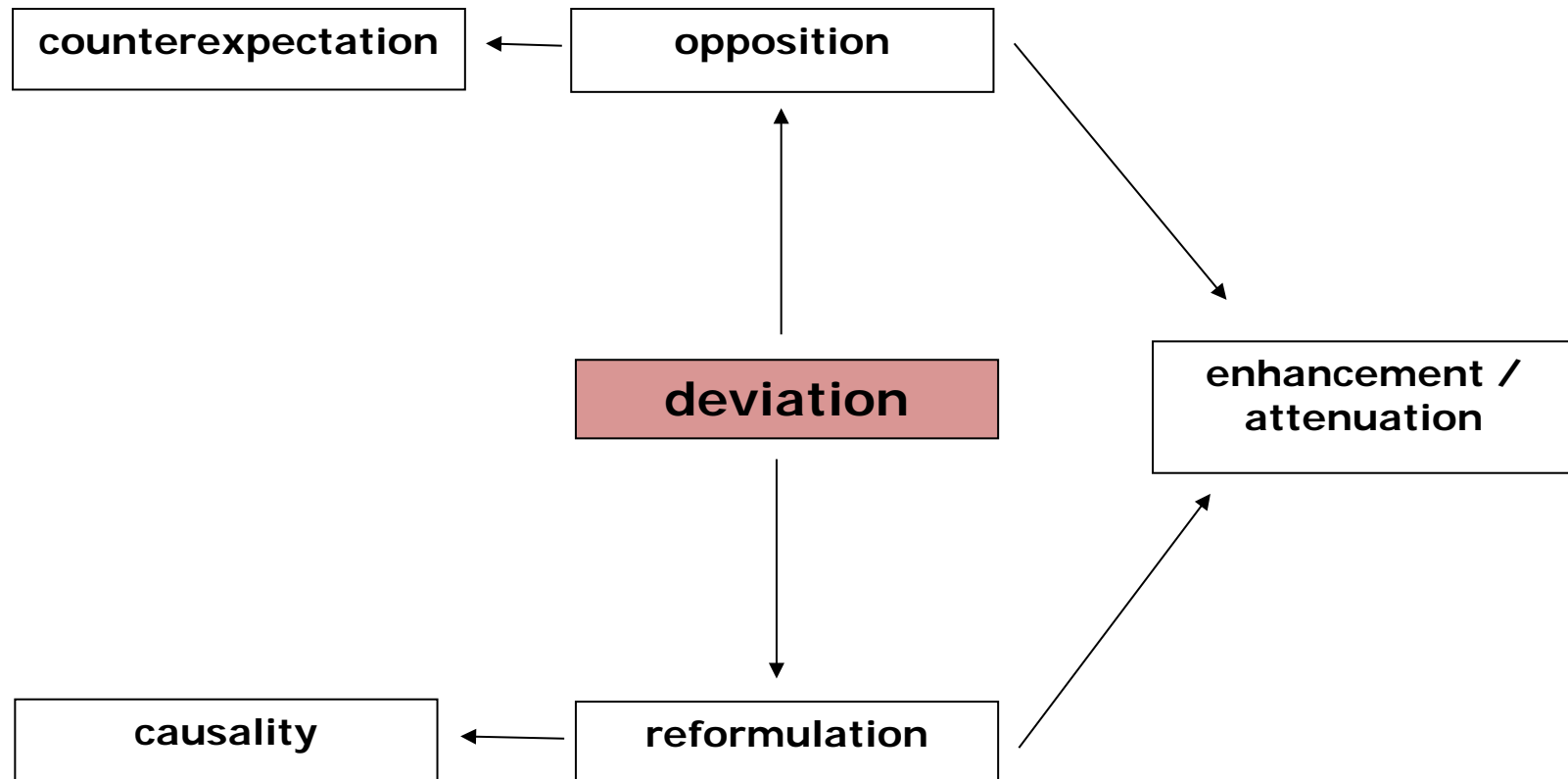


Meaning distribution (2)

- dominant effects:
 - FR vs. DU: *eigenlijk* “precision” and “causality”; *en fait* remarkable shift from even distribution towards “metaperspective” and “precision” in spoken language
 - SP vs. WR: *en fait* “metaperspective”; *eigenlijk* even distribution

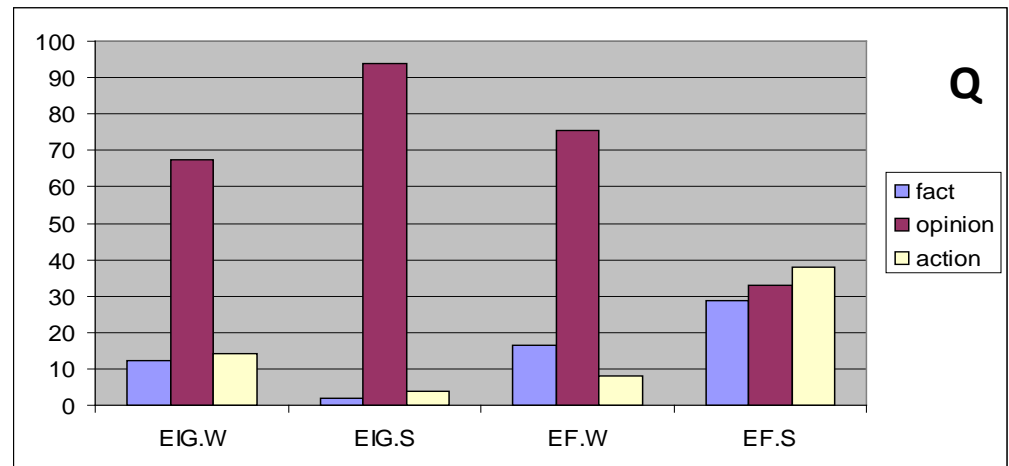
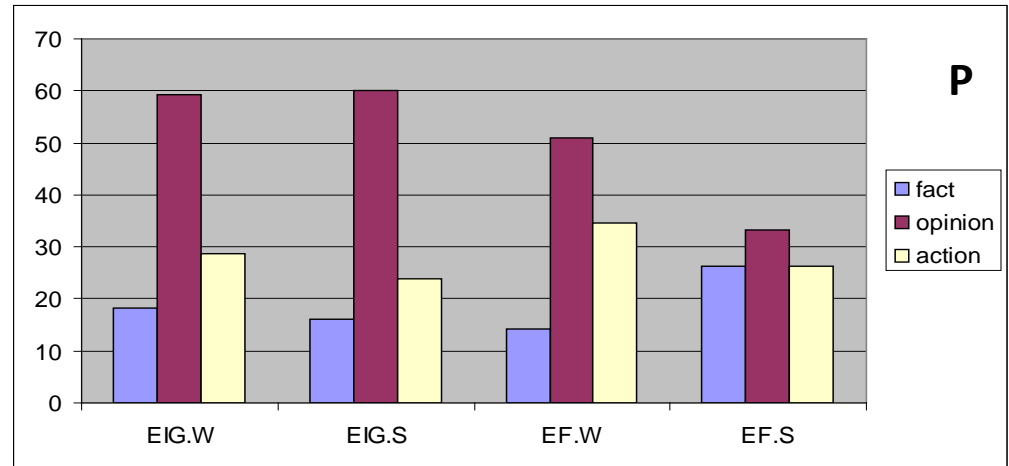


Relational semantic field



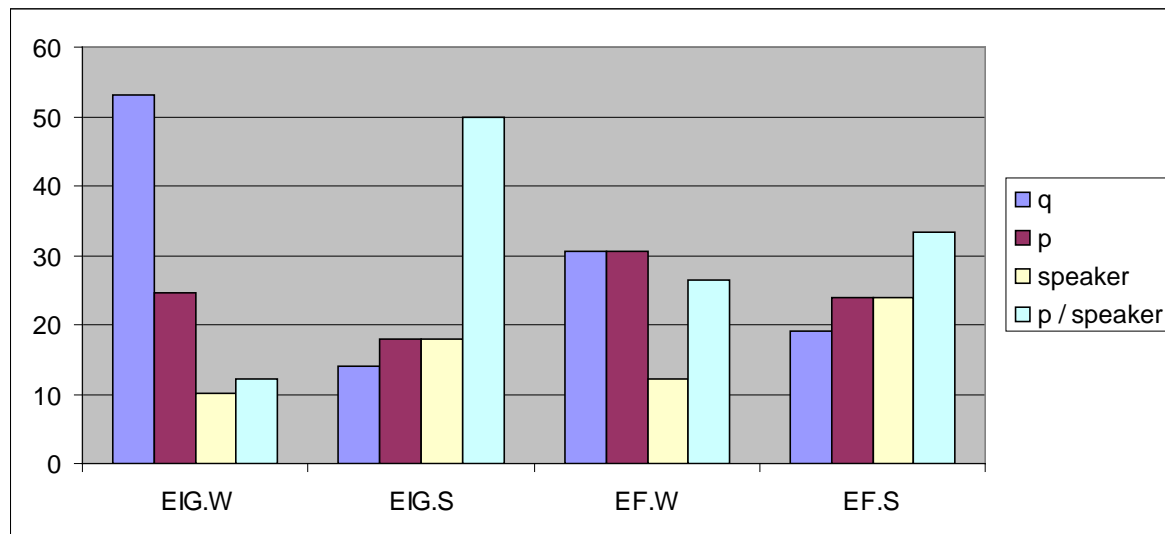
Nature of p and q

- overall dominance of opinions
- *eigenlijk*: increase of opinions from p to q
- *en fait*: lower number of opinions, increase of actions from p to q



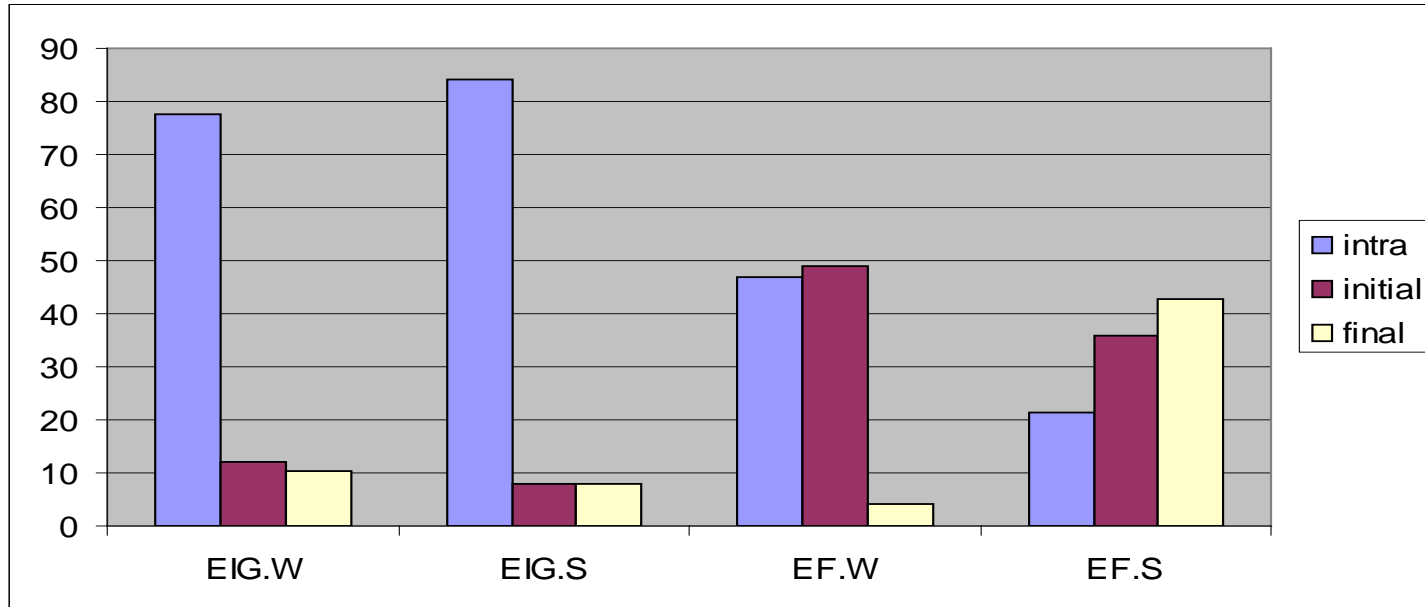
Scope of *en fait* / *eigenlijk*

- dominant scopes:
 - FR vs. DU: *eigenlijk* “local” and “extended speaker”; *en fait* more even distribution
 - SP vs. WR: drop in local scopes to the benefit of extended speaker scopes



Position of *en fait* / *eigenlijk*

- dominant positions
 - FR vs. DU: *en fait* “initial / final”; *eigenlijk* “intra”
 - SP vs. WR: rise of final positions in spoken French; even distribution between registers in Dutch



Results

- **EIGENLIJK:**

- no significant differences between SP and WR for *meaning*, *nature* and *position*
- *scope* is significant (p / speaker ↑), but to be explained by a higher number of pragmatic inferences ‘floating’ around *eigenlijk* in SP
- overall, *eigenlijk* behaves in similar ways in SP and WR
BUT it is much more frequent in SP! (3,1 > 25,93)

- **EN FAIT:**

- significant differences for *meaning* (more DM, metaperspective in SP) and *scope* (speaker, p / speaker)
- somewhat deviant results for *nature* (more actions in SP) and *position* (more final)
- on the whole, however, *en fait* becomes significantly more subjective in SP

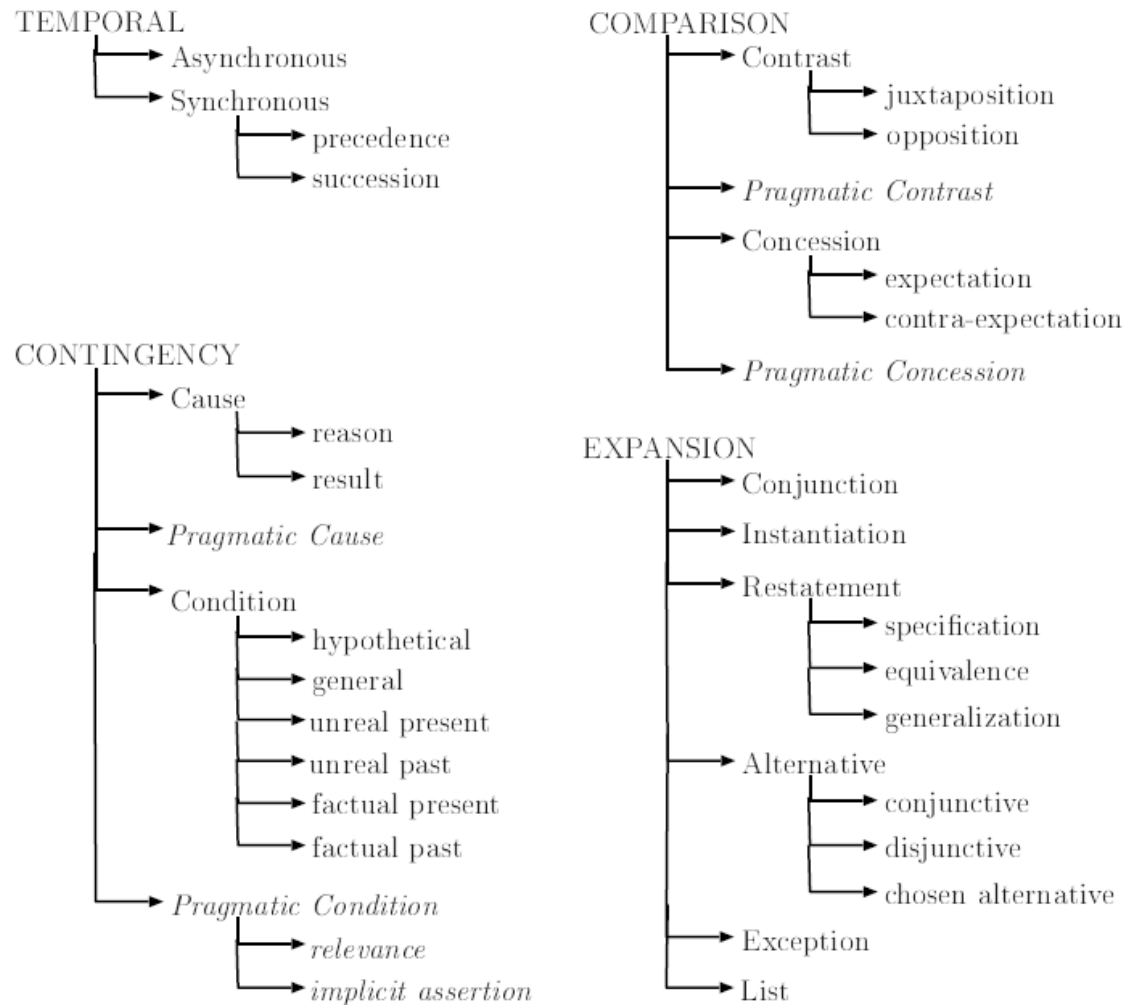
- advantages of a “combined corpus approach”:

- **WR vs. SP:** the status of *en fait* differs according to register
- **FR vs. DU:** not *in feite* but *eigenlijk* is the ‘true’ equivalent of *en fait*

What about annotating DRDs in multilingual data?

- Semantic analysis is prerequisite to corpus annotation ...
- Annotation of what?
 - (Marked) Discourse relations : all vs. specific
 - From coarse-grained to (very) fine-grained ...
- In order to make appropriate cross-linguistic pairings, a common taxonomy of discourse relations must be used, or taxonomy has to be **made** sharable
- **Experiment**: The PDTB taxonomy of discourse relations is theory-neutral, lexically-grounded and has been adapted to many languages.
 - 4 top level types: temporal, comparison, contingency, expansion
 - Can the PDTB taxonomy support multilingual annotations?

PDTB hierarchy of sense tags



Monolingual adaptations of the PDTB taxonomy: an overview

- Use of PDTB with minor modifications
 - Arabic (Al Saif & Markert, 2010)
 - Hindi (Kolachina et al., 2012)
 - Italian (Tonelli et al., 2010)
 - main types from PDTB, but pruning of many subtypes (level 3)
 - a few additions: background (AR), similarity (AR, HI), goal (HI, IT)
 - revision of “pragmatic” tags
- Definition of new, PDTB-style taxonomies
 - Chinese (Huang & Chen, 2011; Zhou & Xue, 2012)
 - Czech (Mladová et al., 2008; Zikánová et al., 2010)
 - French (Danlos et al., 2012)

Multilingual annotation of discourse relations: some questions

1. What are possible types of discourse relations?
2. Are discourse relations types language-independent?
 - Can all relations be encountered in texts from any language?
 - Is there a convergence at least for the Indo-European family?
3. What is the relation between a taxonomy of discourse relations and the connectives of a given language?
4. How can such a taxonomy be used to map discourse connectives from one language to another?

Sense generalization vs. specification

- Abstract goal: extend an existing taxonomy of discourse relations through its application to new languages.
- Consider annotated connectives and their translations.
 - if a connective tagged with sense R appears to be translated by several other target connectives
 - check whether R should not be specified into several subtypes of discourse relations R_1, R_2, \dots, R_n
- Consider fine-grained senses initially defined.
 - if senses R_1, R_2, \dots, R_n are often confused by annotators in the new language, and dispreferred in favor of supersense R
 - check whether only R should be kept in the multilingual taxonomy

Semantic or pragmatic meaning?

- Connectives can convey in context other discourse relations than the one(s) they semantically encode.
 - example: *and* has a core meaning of addition but is often used to convey temporal or causal relations
 - it is not semantically ambiguous (e.g. Carston, 2002, Recanati, 2012), but non-core meanings are constructed by inference
- What level of meaning should be annotated?
 - pragmatic meaning is the one conveyed in context and should be annotated
 - frequency information could determine the distinction between ad-hoc pragmatic meanings and core semantic meanings, if a “dictionary” of connectives must be constructed

A multilingual annotation experiment with the PDTB taxonomy

- Corpus data
 - English (pivot language), French, German, Dutch, Italian
 - parallel corpus gathered from the www.PressEurop.eu website
 - ca. 2500 words/language, 4 texts (different source languages)
- Connectives
 - spotted using PDTB criteria: 54 tokens and 23 types in English
 - annotation of explicitly translated connectives only
 - consequence: up to 50% loss for the annotation!
- Procedure
 - annotators used the PDTB annotation manual
 - two expert annotators per language
 - use double labels for ambiguity ('or') and compositionality ('and')

Monolingual inter-annotator agreement

PDTB level	English	French	German	Dutch	Italian
1	98%	95%	95%	90%	94%
2	67%	69%	71%	60%	63%
3	44%	48%	51%	38%	42%

- **Observations**
 - similar results across languages
 - lower than PDTB results at third level (44% vs. 77% on average)
- **Similar cases of disagreement in all languages**
 - difference between concession/contrast: 50% agreement (40% at level 3)
 - different types of conditionality: 40% agreement
 - use of pragmatic tags: agreement in only 16% of the cases

Multilingual inter-annotator agreement

PDTB level	English/French	English/German	English/Dutch	English/Italian
1	91%	90%	88%	85%
2	67%	66%	64%	58%
3	42%	51%	33%	35%

- **Two distinct phenomena:**
 - at level 1, lower multilingual than monolingual agreement scores
 - at levels 2-3, comparable mono- and multilingual agreement scores
- **Additional multilingual disagreements**
 - all cases of disagreement at level 1 were checked with annotators
 - cause: meaning shifts due to translation (temporal → causal, etc.)
 - about 10% of the cases
 - due to the use of parallel corpora (one is a translation of the other)

Adjustments to the PDTB Taxonomy

1. Temporal

- synchronous
- asynchronous
 - precedence
 - succession

2. Contingency

- cause
- reason
 - pragmatic
 - non-pragmatic
- result
 - pragmatic
 - non-pragmatic
- condition
 - pragmatic
 - non-pragmatic

Generalizations
Specifications

3. Comparison

- contrast
- concession
- pragmatic
- non-pragmatic
- parallel

4. Expansion

- conjunction
- instantiation
- restatement
 - specification
 - equivalence
 - generalization
- alternative
- exception
- list

➔ Similar to previous,
monolingual proposals

Adjustments to the PDTB Taxonomy

- Suppression of problematic tags not useful to distinguish between several connectives.
 - sub-types of conditional and alternative relations
- Addition of tests to help annotators to converge on problematic categories.
 - Substitution test to distinguish between concession and contrast (replacement by *whereas* or *although*).
 - Paraphrase test to distinguish between semantic and pragmatic tags.
- Addition of one tag to account for comparisons without oppositions.
 - COMPARISON: parallel
 - Necessary to annotate: *similarly*, *de même*, etc.

Multilingual annotation experiment with the revised taxonomy

- Corpus data
 - same five parallel languages, again from www.PressEurop.eu
 - ca. 8500 words/language, 10 texts (different source languages)
- Connectives
 - defined using PDTB criteria: 203 tokens and 36 types in English
 - annotation of translated connectives only: from 136 to 155 tokens per language pair
- Procedure
 - one expert annotator per language
 - instructed to use new taxonomy, double tags allowed ('or', 'and')
 - additional explanations for the contrast/concession distinction

Multilingual inter-annotator agreement

PDTB level	English/French	English/German	English/Dutch	English/Italian
1	94%	93%	88%	93%
2	85%	74%	75%	78%
3	75%	66%	69%	66%
4	66%	93%	62.5%	70%

- **Observations**
 - improved inter-annotator agreement at levels 2 and 3
 - disagreements at level 1 still due to meaning shifts in translation
 - adequate coverage of all 36 connective types

Conclusions

- Our experiments have confirmed that the PDTB is operational for multilingual annotations.
 - our adjustments corroborate previous (monolingual) proposals
 - revised taxonomy led to improved agreement
- However, its granularity must also be evaluated.
 - experiments were still English-centered
 - the interchangeability of connectives annotated with the same tags should be checked mono- and multilingually.
- Larger-scale annotations involving more diverse languages should further validate these results.
- Use of spoken data will need additional adjustments (cf. focus on “ideational” meanings)

Conclusions (cont'd)

- Using instructions relying on linguistic rather than metalinguistic judgments increases agreement level.
- Semantic definitions should be accompanied by paraphrases and substitution tests as well as examples.
- Cross-linguistic annotations will have to deal with the problem of implicit relations and alternative lexicalisations.
 - frequent explicitations and implicitations in translations,
 - Are alternative lexicalisations (e.g. gerund, morphemic marker, ...) DRDs?

Take home message

- **Annotation of DRDs** requires semantico-pragmatic analysis in many different languages, taking into account genre and register.
- **Multilingual** annotation requires sound **tertium comparationis** (or rather **tertia comparationis**)
- A **combined corpus analysis** (comparable + parallel) favours construction of sound **tertia comparationis**
- **Inter-annotator agreement** is measure of soundness of the taxonomy/classification/annotation scheme.

References

- Bolly, Catherine T. , Ludivine Crible, Liesbeth Degand et Deniz Uygur-Distexhe (2015). « MDMA. Un modèle pour l'identification et l'annotation des marqueurs discursifs "potentiels" en contexte », *Discours* [En ligne].
- Gast, Volker (2012). Contrastive linguistics: Theories and methods. In Kortmann, B. and J. Kabatek (eds.), *Dictionaries of Linguistics and Communication Science: Linguistic theory and methodology*. Berlin: Mouton de Gruyter.
- Gast, Volker (2011). Slides of LOT-Summerschool course on Contrastive Linguistics
- König, Ekkehard. 2012. « Contrastive linguistics and language comparison ». *Languages in Contrast* 12 (1): 3-26.
- Malchukov, A. 2004. "Towards a semantic typology of adversative and contrast marking". *Journal of Semantics*, 21 (2), 177-198
- Mortier, L. & Degand, L. (2009). Adversative discourse markers in contrast: the need for a combined corpus approach. *International Journal of Corpus Linguistics* 14(3), 338-366.
- Pander Maat, H. & Degand, L. 2001. "Scaling causal relations and connectives in terms of Speaker Involvement". *Cognitive Linguistics*, 12 (3), 211-245

References

- The PDTB Research Group. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. IRCS Technical Reports Series, 99p.
- Schwenter, S. 2000. "Viewpoints and polysemy: Linking adversative and causal meanings of discourse markers". In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause–Condition–Concession–Contrast. Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 257-282.
- Van der Auwera, Johan. 2012. « From contrastive linguistics to linguistic typology ». *Languages in Contrast* 12 (1): 69-86.
- Webber B. & Joshi A. (2012). Discourse structure and computation: Past, present and future. *Proceedings of the ACL-2012*, Jeju, Republic of Korea, 42–54.
- Verhagen, A. 2000. "Concession implies causality, though in some other space". In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause–Condition–Concession–Contrast: Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 361-380
- Zufferey, Sandrine, Liesbeth Degand, Andrei Popescu-Belis & Ted Sanders (2012). Empirical validations of multilingual annotation schemes for discourse relations. *Proceedings of the 8th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*. Pisa, 3-5 October 2012, 77-84.[Available online at: http://sigsem.uvt.nl/isa8/ISA-8_proceedings.pdf
- Zufferey, S. & Degand, L. (in press). Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguistics and Linguistic Theory*.