



**Agenda  
Management Committee Meeting**

**COST Action no. IS1312**

**Action Title: Structuring Discourse in Multilingual Europe (TextLink)**

**Venue: Room 24, 25 Dózsa György Street, Budapest**

**Country: Hungary**

**Date: 11 April 2016 (11.00 – 13.00)**

1. Welcome (Action Chair)
2. Adoption of agenda (Action Chair)
3. Approval of minutes and matters arising from last meeting (Action Chair)  
*Paper TL201604.01*
4. Update from the Action Chair (Action Chair)
5. Promotion of gender balance, Early Stage Researchers (ESRs), inclusiveness  
(Action Chair)  
*Memorandum for the record*
6. Update from COST Association (Action Chair)
7. Budget (Grant Holder Manager)  
*Paper TL201604.02*
8. Grant Period 2
  - a. Website and dissemination update (Dissemination Manager)
  - b. Training School Valencia (Training School Coordinator)  
*Paper TL201604.03*
  - c. STSM status (Training School Coordinator)
  - d. Dispol meeting
  - e. Work achieved this period (WG leaders)
9. Work plan for period 3 (Steering Committee)
10. AOB
11. Location and date of next meeting
12. Summary of MC decisions taken (Grant Holder Manager)
13. Closing

*Appendix TL.201604.01:  
Summary of Actions and Decisions taken from Minutes of  
last MC Meeting (January 2015)*

### **Summary of MC decisions taken**

- The minutes of the previous meeting, as well as a list of MC decisions taken since the last meeting, were approved.
- The MC approved a maximum travel reimbursement rate of 400€ (Fribourg meeting).
- The MC approved the training school proposal (see appendix).
- The MC approved the membership of Lithuania and Slovenia.
- The MC approved Brazil's membership to the Action, as observers.

### **Actions required**

- GH manager to clarify with COST Association whether TextLink still falls under ISCH.
- WG2 and WG3 co-leaders to decide how many participants will be invited to Fribourg.
- Training School Coordinator to receive suggestions for specific training topics and trainers from WG co-leaders, and address other pending questions.
- GH manager to clarify with COST Association whether master's students can be reimbursed for participation at training schools.
- GH manager, dissemination manager and Piotr Pezik (PL) to explore resources that will allow sharing of member bibliographies.
- GH manager to collect conference proceedings.
- Action Chair to communicate to Lithuania and Slovenia, as well as the COST Association, the MC's decision to approve these countries as members of the Action.
- Action Chair to communicate this decision to accept Brazil as an observer to the Action to the COST Association, and to begin the process of accepting Brazil into the Action.
- GH manager to check with COST Association whether official reorganization of WGs is possible.

*Appendix TL.201604.02:  
Budget and expenses for Grant Period 2*

**TextLink: Period 2 Budget and Current Expenditure (as at  
25/03/2016**

	Work plan (July 2016)	Current Expenditure	Foreseen Expenditure
<b>Meetings</b>			
Action conference (Budapest) <i>3.5d, 78 participants, 78 reimbursed</i>	40,200 €	0 €	67,500 €
Spoken Discourse (Saarbrücken) <i>2d, 42 participants, 20 reimbursed</i>	13,200 €	13,067 €	0 €
WG1 / WG4 meeting (Warsaw) <i>2d, 16 participants, 15 reimbursed</i>	17,300 €	8,759 €	0 €
<b>Dissemination conferences</b>			
Dissem. Portugal	1,120 €	890 €	0 €
Dissem. Canada	1,910 € cancelled		0 €
<b>Training schools</b>			
Training School (Valencia) <i>5d, 12 trainers, 38 trainees (28 reimbursed)</i>	43,640 €	34,481 €	0 €
<b>Dissemination</b>			
Website costs	1,500 €	1,500 €	0 €
<b>STSMs</b>			
STSMs <i>10 planned, 8 awarded, 2 cancelled</i>	18,000 €	6,860 €	3,500 €
<b>OERSA</b>			
Banking costs	520 €	0 €	520 €
<b>FSAC</b>			
15% of total scientific	20,609 €	9,834 €	10,728 €
<b>TOTAL</b>			
<b>Total</b>	<b>157,999 €</b>	<b>75,390 €</b>	<b>82,248 €</b>

*Appendix TL.201604.03:  
Valencia training School*



## REPORT

# **Methods and tools for the analysis of discourse relational devices**

1<sup>st</sup> TextLink Training School

COST Action IS1312

TextLink:

Structuring Discourse in Multilingual Europe

**València (Spain), 18<sup>th</sup>-22<sup>nd</sup> January, 2016**





## Meeting Details

**Host Institution:** University of Valencia

**Location:** Valencia (Spain), Facultat de Filologia, Av. Blasco Ibáñez 32

**Dates:** January, 18-22, 2016.

**TextLink Training School officer:** Maria Josep Cuenca

**Organizing committee coordinators:** Maria Josep Cuenca & Salvador Pons

## Aims

- Delivering intensive training on theoretical and empirical approaches to the description, identification and annotation of DRDs in multiple languages.
- Providing a theoretical and empirical introduction to discourse coherence with a focus on DRDs.
- Training in methods and tools for the identification and annotation of DRDs in authentic data, including statistical methods for data analysis.
- Improving trainer-trainee collaboration, fostering of ideas for future collaboration.

## Structure

### General sessions

Topic 1 (2 sessions): Coherence relations and DRD identification

Topic 2 (3 sessions): Cross-linguistic variation

Topic 3 (3 sessions): Corpus research, annotation theories and tools

Topic 4 (2 sessions): Dictionaries/Lexicons

### Lab sessions

Hands-on sessions focusing on technical aspects and statistics. Specific topics: how to assemble a sound corpus, reliability, annotator agreement, experimental design, hypotheses, analysis of results, and so on.

### Research speed dating session (see instructions in Annex 3)

Trainers and trainees talk for a short period of time about their current research or research interests. Meet up with other researchers in the same field, discuss crucial questions and methodology, share your insights and experiences. Trainees must be prepared to summarize their own research in 5 minutes.

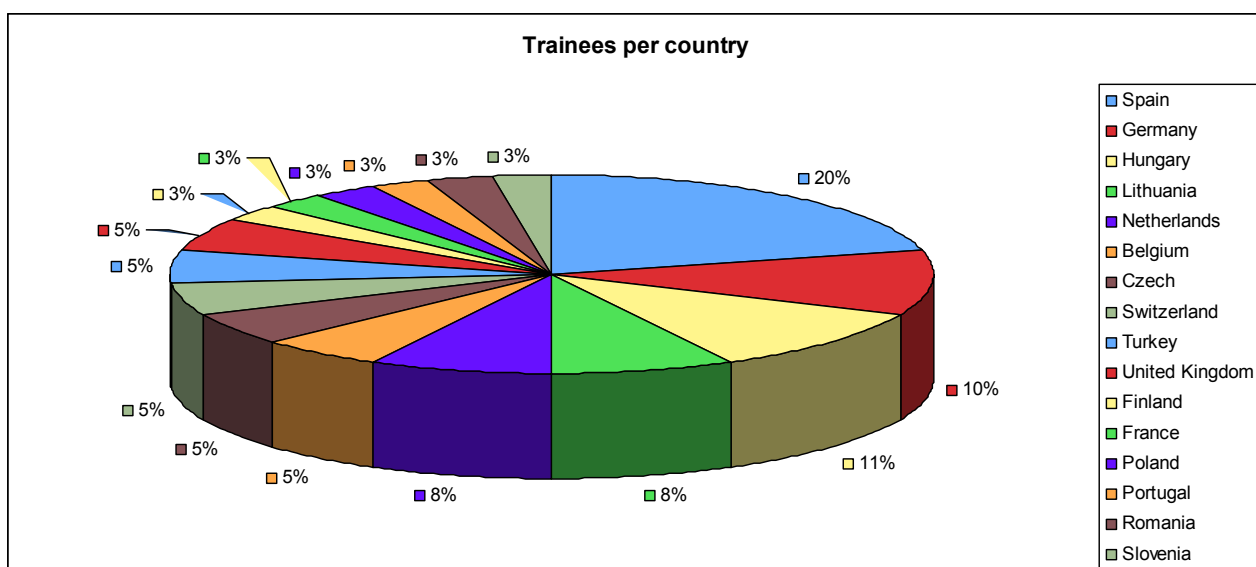
### Side Event (see programme in Annex 6)

“Discourse Relational Devices”, 3<sup>rd</sup> International Conference on Linguistic & Psycholinguistic Approaches to Text Structuring (LPTS 2016), Universitat de València, January, 24-26, 2016.

## Participants

An application form was distributed among TextLink members as an initial step (see form in Annex 6). 42 applications were received and 6 more people were on a waiting list. The final number of participants is as follows (see list of participants, Annex 1):

- 38 trainees (27 grants)
- 12 trainers



Some of the TextLink TS Participants also participated in the side event “Discourse Relational Devices”, *3<sup>rd</sup> International Conference on Linguistic & Psycholinguistic Approaches to Text Structuring* (LPTS 2016, Universitat de València, January, 24-26):

- 25 trainees
- 5 trainers

## Schedule

**Registration: Monday 18, 8:15-8:40**

**Opening ceremony: Monday 18, 8:40-9:00**

	<b>General session Room 406</b>	<b>General session Room 406</b>	<b>Lab session Room 406</b>
	9-10.30	11-13.30	15.00-18
<b>Mon 18</b>	Coherence relations and DRD identification: theory and analysis T. Sanders / W. Spooren	Annotation theories and tools A. Nedoluzhko	Lab: Corpus research: Methodology and statistics W. Spooren/ T. Sanders
<b>Tues 19</b>	Coherence relations and DRDs identification: converging evidence T. Sanders / W. Spooren	Corpus exploration of discourse relations in PDT 3.0 and PDTB J. Mírovský	Lab: Annotation tools A. Nedoluzhko
<b>Wed 20</b>	Cross-linguistic variation of DRDs J. Visconti	Corpus exploration of discourse relations in RST M. IruSKIETA	<i>Research 'speed dating'</i> (Room: Espai cultural, 1 <sup>st</sup> floor)
<b>Thur 21</b>	Cross-linguistic variation: DRDs identification and annotation L. Degand/ S. Zufferey	Building discourse relational device lexicons L. Danlos	Lab: Working with cross-linguistic data S. Zufferey/L. Degand
<b>Fri 22</b>	Typology and DRDs V. Gast	Methodological issues on DRDs dictionary construction: The case of the DPDE S. Pons	Lab: Machine translation techniques to induce multilingual lexica of discourse markers D. Martin de Matos

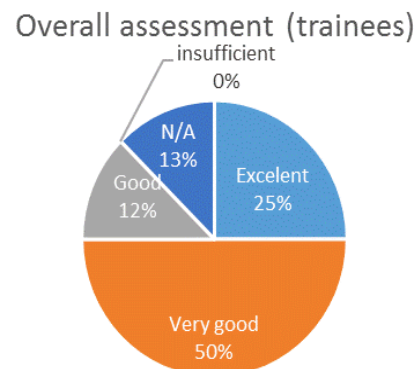
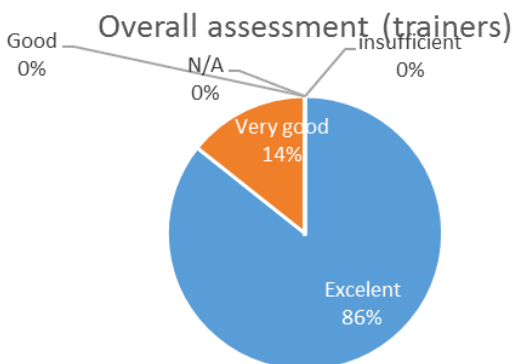
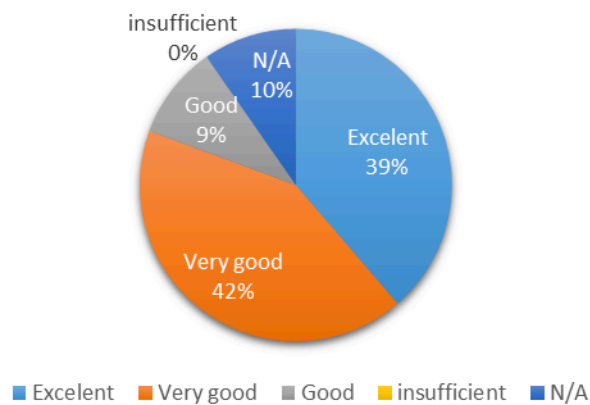
**See session descriptions in Annex 2**



## Evaluation and proposals for the 2<sup>nd</sup> TextLink Training School

By the end of the TS, an evaluation form was given to participants (see Annex xx). 31 evaluations forms were collected, 7 out of 12 filled in by trainers and 24 out of 38 filled in by trainees. The general assessment is excellent or very good, especially among trainers (more than 80%).

### Overall assessment



### 1. General comments about the 1st Textlink TS

Both trainers and trainees highlight the good organisation and the great overall experience, as well as the friendly environment encouraging conversation and academic exchange. From the academic point of view, trainers praise the variety

of topics covered and trainees stress the interest of the topics and the expertise of the trainers.

## 2. Strong points

In line with the general comments, the good organisation of the whole Training School is mentioned by 5 out of 6 trainers and by 15 out of 24 trainees. Although some clarifications are often made later, most participants also highlight the quality and interest of the topics selected and mention the expertise, openness and generosity of the trainers, as well as their motivation.

Both trainers and trainees seem to evaluate positively or very positively the research speed dating session. Some of them have even given some tips to improve the experience (some suggest that, if groups of four are created, each person should speak only once, for ten minutes, and then receive feedback from the trainers and the other two trainees). Only two trainees and one trainer express objections about Research Speed Dating activity, mainly on the basis of it being a tough experience for shy people, or being useless due to the short time to explain the research. Overall, it has been considered a success and the organisation has been commended for the results.

Despite some brief remarks commented on in Sections 3 and 4 below, trainees seem to agree on the balance between theory and practice (although the practice is better considered in general) and, with a few exceptions, they do appreciate the different theoretical backgrounds of both the trainers and their fellow trainees, a circumstance that is said to enrich discussion and to create networks and promote joint work.

One of the most frequent items considered a strong point was the presence of lab sessions. Hands-on experience with software and tools was much appreciated.

## 3. Aspects to be improved

The main objection to the 1<sup>st</sup> Training School format has to do with the distribution of the sessions and the time devoted to them. As for the sessions, some trainees consider the second session in the morning extremely long and would have preferred two equally long sessions. Another frequent claim is to have one or two free morning/afternoons. The schedule (from 09:00 to 18:00) is considered too dense. Some

Some students point out that background readings should have been made available earlier for them to have time to get acquainted with the theories presented. They also ask for the slides to be uploaded in advance, and they ask the

trainers to use bigger fonts in their Power Points (and the organisers to get bigger screens, too).

The fact of working in teams as a way to practise theories has been presented as a strong point, but a number of students suggest the need to establish a common aim, a goal-oriented task where each subgroup would have to deliver some kind of application or outcome.

Trainers should have more information about the trainees' background.

Trainers sometimes tried to cover too many aspects. They should focus on some of them.

#### 4. Suggestions for the 2<sup>nd</sup> TextLink TS

From the theoretical point of view, a frequent claim is the need for statistics sessions. Many students even suggest that the organisation should devote a whole day to this particular topic, approaching statistics both theoretically (i.e. providing some very basic notions on the tests, concepts, etc.) and practically, as it was done during the first lab session with SPSS.

A comparison between theories would be welcome, as well as the introduction of theories such as SDRT. A more thorough work with the tools is also asked for. Some students complain that lessons like the PDTB or Graph Anno were taught as regular lectures despite being strongly tool-based, and therefore the students could not follow the explanation as well as they would have liked to.

The languages involved have also been mentioned occasionally: students suggest to strengthen cross-linguistic approaches and take lesser spoken languages into account, especially non-related ones.

The different profiles of the trainees should be taken into account more specifically. Some sessions were difficult for some trainees and some others too basic for other trainees. In fact, the background and particular interests of the students result in two opposite solutions: half of them prefer to adopt a more general, cross-theoretical approach to get a general overview, whereas the other half considers that too theoretical or too overarching lectures result in shallow approaches and prefer a narrower focus, on a few topics, but more in-depth.

Suggested topics and activities:

- detailed analyses of tools and their evaluation together,
- introduction of two tools and a lot of hands-on experience, practical tasks with trainees' own outcomes,

- more hands-on experience on annotation; language-specific annotation after joint (English) in-class annotation,
- discourse segmentation; signalling vs DRD annotation: problems and challenges; understanding different theories: SDRT, RST, PDTB... and relations; applications of discourse annotation: question answering, summarization, sentiment analysis,
- automatic detection of DRDs; automatic parsing of discourse relations; how discourse annotation contributes to other NLP tasks,
- more on psycholinguistic validation,
- properties of individual connectives from languages such as Turkish, Hungarian, Finnish (e.g. compared to English); session on CDA (a really basic one) focused on DRDs,
- summary of the research in some subfield (e.g. causal relations from cognitive perspective in Italian),
- lessons on spoken data (theory and tools),
- lessons on statistics (a full day devoted to them); step by step guide and hands on tutorial obtaining frequency counts,
- corpus linguistics methods,
- a summary session taking into account the findings of this COST action.

## Annex 1. List of Participants

### Trainers

1.	DANLOS Laurence	<i>laurence.danlos@inria.fr</i>	FR
2.	DEGAND Liesbeth	<i>liesbeth.degand@uclouvain.be</i>	BE
3.	GAST Volker	<i>volker.gast@uni-jena.de</i>	DE
4.	IRUSKIETA Mikel	<i>mikel.iruskieta@ehu.eus</i>	ES
5.	MARTINS DE MATOS David	<i>david.matos@inesc-id.pt</i>	PT
6.	MIROVSKY Jiri	<i>mirovsky@ufal.mff.cuni.cz</i>	CZ
7.	NEDOLUZHKO Anna	<i>nedoluzko@ufal.mff.cuni.cz</i>	CZ
8.	PONS BORDERIA Salvador	<i>salvador.Pons@uv.es</i>	ES
9.	SANDERS Ted	<i>T.J.M.Sanders@uu.nl</i>	NL
10.	SPOOREN Wilbert	<i>w.spooren@let.ru.nl</i>	NL
11.	VISCONTI Jacqueline	<i>j.visconti@unige.it</i>	IT
12.	ZUFFEREY Sandrine	<i>sandrine.zufferey@gmail.com</i>	CH

### Trainees

1.	ABUCZKI Agnes	<i>abuczki.agnes@gmail.com</i>	HU
2.	ALBOM Ana	<i>ana.albom@unibas.ch</i>	CH
3.	ALBELDA, Marta	<i>marta.albelda@uv.es</i>	ES
4.	ALBUQUERQUE F Robson	<i>albuquerque@zedat.fu-berlin.de</i>	DE
5.	ANTUNES Sandra	<i>sandra.antunes@clul.ul.pt</i>	PT
6.	BLYTHING Liam	<i>l.blything@lancaster.ac.uk</i>	UK
7.	CIMMINO Doriana	<i>doriana.cimmino@unibas.ch</i>	CH
8.	DOBROVOLJC Kaja	<i>kaja.dobrovolic@trojina.si</i>	SI
9.	DUPONT Maite	<i>maite.dupont@uclouvain.be</i>	BE
10.	ESTELLÉS, Marla	<i>maria.estelles@uv.es</i>	ES
11.	FODOR Alexandra	<i>sanfodor@gmail.com</i>	HU
12.	FURKO Balint Peter	<i>furko.peter@gmail.com</i>	HU
13.	GABARRO-LOPEZ Silvia	<i>silvia.gabarro@unamur.be</i>	BE
14.	GRZECH Karolina	<i>298351@soas.ac.uk</i>	UK
15.	HOEK Jet	<i>j.hoek@uu.nl</i>	NL
16.	JINOVA Pavlina	<i>jinova@ufal.mff.cuni.cz</i>	CZ
17.	KRIELKE Pauline	<i>pauline.krielke@gmail.com</i>	DE
18.	KURFALI Murathan	<i>murathankurfali@gmail.com</i>	TR
19.	LAITINEN Janne	<i>janne.v.laitinen@helsinki.fi</i>	FI
20.	LAVID Julia	<i>julavid@filol.ucm.es</i>	ES
21.	LLOPIS Ana	<i>ana.llopis@ucv.es</i>	ES
22.	MAZEIKIENE Viktorija	<i>viktorija.mazeikiene@gmail.com</i>	LT
23.	MESTRE MELIA Marti	<i>marmesme@alumni.uv.es</i>	ES
24.	NAGY Anna	<i>nagyanna.de@gmail.com</i>	HU
25.	OGRODNICZUK Maciej	<i>maciej.ogrodniczuk@jpipan.waw.pl</i>	PL
26.	PASCUAL ALIAGA Elena	<i>Elena.Pascual@uv.es</i>	ES
27.	POLAKOVA Lucie	<i>polakova@ufal.mff.cuni.cz</i>	CZ
28.	POSTOLEA Sorina	<i>sorinapostolea@gmail.com</i>	RO
29.	SALAMEH JIMENEZ Shima	<i>salamehjimenez@gmail.com</i>	ES
30.	SANCHEZ-CARDENAS Beatriz	<i>bsc@ugr.es</i>	ES
31.	SANTANA COVARRUBIAS Andrea	<i>A.C.SantanaCovarrubias@uu.nl</i>	NL
32.	SAYGINER Ege	<i>egesay@gmail.com</i>	TR
33.	SCHOLMAN Merel	<i>m.c.j.scholman@coli.uni-saarland.de</i>	DE
34.	SLIOGERIENE Jolita	<i>j.sliogeriene@mruni.eu</i>	LT
35.	TIMOTHEE Bernard	<i>timothee.bernard@ens-lyon.org</i>	FR
36.	UCLÉS RAMADA Gloria	<i>Gloria.ucles@gmail.com</i>	ES
37.	VALUNAITE-OLESKVICIENE Giedre	<i>gentrygiedre@gmail.com</i>	LT
38.	WEI Yipu	<i>y.wei1@uu.nl</i>	NL



## Annex 2. Sessions description

### Coherence relations and DRD identification: theory and analysis

Trainers: Ted Sanders (Universiteit Utrecht) / Wilbert Spooren (Radboud Universiteit Nijmegen)

Language users communicate through discourse. The constituting property of discourse is that it shows coherence: people make a coherent mental representation of the information in the discourse. The discourse itself contains (more or less) overt signals that direct this interpretation process, among them discourse relational devices (DRDs) like connectives and cue phrases.

We focus on *coherence relations* that establish the relationship between discourse segments, such as *Cause-Consequence* and *Contrast*. These relations are conceptual and they can, but need not, be made explicit by DRDs (*because, so, however, although*) and lexical cue phrases (*For that reason, As a result, On the other hand*).

Some dominating accounts of coherence relations are introduced, compared and discussed. We will discuss their use in discourse annotation, as well as their theoretical backgrounds. Special attention will be paid to underlying dimensions on which various accounts converge. In the Lab session, we will work on concrete issues of annotation of corpus case.

### Annotation theories and tools

Trainer: Anna Nedoluzhko (ÚFAL, Charles University Prague)

Discourse coherence is a complex natural language phenomenon which is achieved by different linguistic means (e.g., anaphoricity, information structure, discourse markers and connectives, rhetorical structure of text, etc.). Many approaches in computational linguistics are used to capture discourse relations and find practical applications.

This session focuses on discussing theories and approaches applied to the annotation of discourse phenomena in different languages, such as Rhetorical Structure Theory, Penn Discourse Treebank, Segmented Discourse Representation Theory and so on. Participants will have the opportunity to compare these approaches to see which phenomena are central to them and which ones are less prominent. We will also introduce the tools of discourse annotation and demonstrate esp. TrEd, PDTB and MMAX2.

### Corpus research: Methodology and statistics

Trainers: Wilbert Spooren (Radboud Universiteit Nijmegen) / Ted Sanders (Universiteit Utrecht)

This lab session focuses on corpus research at the discourse level: How can you assemble a corpus to do your investigations on, in a valid and reliable way? How can you search in existing corpora and what is best format to annotate and analyze cases? Participants discuss concrete cases taken from corpora of language use in various languages. We analyze coherence relations and DRDs. Annotation Tools – some of them automated – are introduced and put to use. The next methodological step is to

determine whether several analysts agree: interrater reliability. Methods to compute this are introduced and participants will use them during the session. Further methodological implications are discussed. Finally, we discuss statistical methods (in SPSS or R) that help answer research questions, such as: do DRDs behave differently in one genre than in another?

### **Coherence relations and DRD identification: converging evidence**

Trainers: Ted Sanders (Universiteit Utrecht) / Wilbert Spooren (Radboud Universiteit Nijmegen)

In a cognitive approach to coherence relations, it is important to account for the relationship between discourse as a linguistic object and the mental representation people have or make of it. Such an approach requires an interdisciplinary methodology of converging evidence. We discuss studies using both linguistic and psycholinguistic research methods and data, varying from text analysis to on-line discourse processing and language acquisition. Finally, we explore the relationship between discourse coherence and genre.

### **Corpus exploration of discourse relations in PDT 3.0 and PDTB**

Trainer: Jiří Mírovský (Charles University Prague)

We will explore discourse relations in the Prague Dependency Treebank and in the Penn Discourse Treebank using the PML-Tree Query system, a general and powerful system for querying treebanks. We will learn basics of the query language on syntactic trees and use it later for searching for discourse relations. We will show that as a result of a query search, we can either get a sequence of individual occurrences of the query pattern in the data, or a summary of the occurrences in the whole data defined by a system of output filters.

### **Lab: Annotation tools**

Trainer: Anna Nedoluzhko (ÚFAL, Charles University Prague)

Following the general session on “Annotation theories and tools”, trainees will be invited to use several annotation tools.

### **Cross-linguistic variation on DRD**

Trainer: Jacqueline Visconti (University of Genoa)

Given the multifunctionality and context-boundness of DRDs, linguistic variation is a tricky question. The first part of this interactive lecture will focus on central theoretical issues in contrastive analysis, such as the notion of tertium comparationis, the balance between monolingual and comparative methods, the role of corpora in contrastive studies. In the second part, a selection of case-studies of contrastive investigations on DRDs in mostly European languages will be highlighted and discussed. methods will rely on multilingual corpora, such as Eurparl or VoxEurop, from which translation equivalents will be elicited as empirical data.

## Corpus exploration of discourse relations in RST

Trainer: Mikel Iruskieta (University of the Basque Country)

In the RST framework, there are several discourse-annotated corpora available in individual languages, such as:

English <<https://catalog ldc.upenn.edu/LDC2002T07>>,

Spanish <<http://corpus.iingen.unam.mx/rst/>>,

Brasilian Portuguese <<http://www.nilc.icmc.usp.br/nilc/tools/corpora.htm>>, German

<<https://www.ling.uni-potsdam.de/acl-lab/Forsch/pcc/pcc.html>>, and

<<http://ixa2.si.ehu.es/diskurtsoa/en/>>, among others.

Some of them can be consulted and several tools have been developed for corpus exploration (exploration honekin zalantza, baina literaturan ikusi baduzu, aurrera!).

There is also a small multilingual aligned RST corpus <<http://ixa2.si.ehu.es/rst/>>, which can be explored for getting information about different linguistic phenomena.

After the annotation process is over, evaluation is necessary to check reliability (precision and recall). In order to do so, a sound evaluation method and some search tools (which can be used in multilingual corpus) were developed: i) to study whether the annotators were consistent when looking for the relations or signals in a kwic style, ii) to check the aligned segments in different languages, iii) to check a kind of macrostructure of RS-tree looking for the RST relations that are linked to the most salient unit, and iv) to look for any information in the corpus based on part of speech.

In this session, I will present this method and the tools developed to consult the multilingual RST treebank we have developed in the Ixa research group at the University of the Basque Country.

## Cross-linguistic variation: DRD identification and annotation

Trainers: Liesbeth Degand (Université catholique de Louvain) / Sandrine Zufferey (University of Fribourg)

While DRDs are found in (probably) all languages of the world, important variations exist in the number of DRDs languages display to express a given relation, even between typologically related languages. An overview will be given of the types of variation that exist along different types of dimensions: form vs. function, semasiology vs. onomasiology, and how they can be empirically investigated on the basis of different types of data: parallel (translation, including fiction, non-fiction, subtitles, ...) or comparable (same text types in different languages), both in speech and writing.

## Building discourse relational device lexicons

Trainer: Laurence Danlos (Université Paris Diderot)

Discourse relational devices (DRDs) are (simple or compounds) lexical items that express discourse relations between two discourse segments. For French we developed a lexicon of DRDs which records for each entry its syntactic category and its sense(s) (i.e. which discourse relation(s) it expresses) along with possible other information (e.g. constraint on its position) and examples. A first version, developed

from linguistic knowledge, was revised after a discourse annotation experiment, and we will present the two methods. DRDs lexicons exist for other languages, German for example. We will make a comparison of the two resources with a reflection on the methods that was used to build them.

### **Lab: Working with cross-linguistic data**

Trainers: Sandrine Zufferey (University of Fribourg) / Liesbeth Degand (Université catholique de Louvain)

Following the general session on “Cross-linguistic variation: DRD identification and annotation”, trainees will be invited to participate in a multilingual annotation experiment.

### **Typology and DRD**

Trainer: Volker Gast (University of Jena)

The cross-linguistic, corpus-based study of discourse relational devices requires corpora annotated at various levels, minimally syntax and semantics. In this session I will demonstrate how we can create a cross-linguistic sample of sentences annotated at two levels by carrying out the following steps: (i) preparation of the data (e.g. syntactic parsing of a sample), (ii) manual corrections, (iii) enriching the sample with lexical-semantic information, and (iv) manually annotating it with sentence-semantic and pragmatic information.

### **Methodological issues on DRDs dictionary construction: The case of the DPDE**

Trainer: Salvador Pons Bordería (Universitat de València)

Building a dictionary of DRDs seems a paradoxical activity, given the functional explanation assigned to them. Indeed, the lexicographical description of DRDs implies challenges both for the lexicographer and for the pragmatist; the former has to review received wisdom related to what counts as a definition, synonymy or even meaning. The latter has to transform an onomasiological approach, based on functions, into a semasiological approach, based on forms.

This session will reflect on the issues above on the experience from the *Diccionario de Partículas Discursivas del Español* (Briz, Pons and Portolés, online since 2003) ([www.dpde.es](http://www.dpde.es)).

### **Lab: Machine Translation Techniques to Induce Multilingual Lexica of Discourse Markers**

Trainer: David Martin de Matos (University of Lisbon)

Discourse markers are universal linguistic events subject to language variation. This lab's work contemplates new methods and approaches for the description, classification, and annotation of discourse markers in the specific domain of the Europarl corpus. The study of discourse markers in the context of translation is crucial due to the idiomatic nature of these structures. Multilingual lexica together with the functional analysis of such structures are useful tools for the hard task of translating

discourse markers into possible equivalents from one language to another. Using Daniel Marcu's validated discourse markers for English, extracted from the Brown Corpus, our purpose is to build multilingual lexica of discourse markers for other languages, based on machine translation techniques.

### Annex 3. Research speed dating instructions

The aim of this activity is to bring young researchers in contact with senior scholars and other trainees in order to obtain feedback and valuable insight on their particular research from a variety of theoretical perspectives.

The activity will run as follows:

- 1) Twelve groups including 1 trainer and 3 students will be created. Each trainer will be randomly associated to one of these groups.
- 2) All students will join one of these groups. No student will remain without a group.
- 3) Each student will have **five** minutes to explain his/her research in an “elevator pitch” way, followed by five more minutes during which the senior scholar will provide his/her feedback. Students not exposing to a trainer can talk to each other and present their respective research to other trainees.
- 4) One of our collaborators will ring a bell in the room: one single ring means ‘five minutes’, which is an orientative time for the student to finish his/her presentation and start receiving feedback; two rings mean ‘ten minutes’ and therefore ‘change of speaker’. When all students have talked to their group trainer, they will change group. **It is vital that the times and changes are respected for the activity to run successfully.**
- 5) After 60 minutes, two rounds of pitches have been completed. A 30’ break will follow, during which trainees can informally meet other trainers and trainees.
- 6) After the break, the groups will re-arrange: every student will join different groups for the next 30+30 minutes. By the end of this second session (round 4), every student will have met four trainers and 8 trainees in 4 different groups.
- 7) The speed-dating session will finish with a second break, during which trainers and trainees meet informally while having a drink.

Time schedule:

Round 1	15:00- 15:30, 12 groups (1 group = 3 students + 1 trainer)
Round 2	15:30- 16:00, 12 groups (1 group = 3 students + 1 trainer)
Break 1	16:00- 16:30
Round 3	16:30- 17:00, 12 groups (1 group = 3 students + 1 trainer)
Round 4	17:00- 17:30, 12 groups (1 group = 3 students + 1 trainer)
Break 2	17:30 -18:00

The organising committee assumes that all of the trainees participate in the ‘research speed dating’ activity. In case you do not wish to join us, please let us know by contacting our staff in the reception desk by Tuesday 19th.

## Annex 4. Social Program

### » Monday 18<sup>th</sup> January, 19-21h: City tour

Valencia's Old town tour. We will visit some of the most important places in the old town.

### » Tuesday 19<sup>th</sup> January, 19.15: Concert in Palau de la Música

**Program:** pieces by Mozart, Ravel and Schubert performed by Herold quartet (Sociedad Filarmónica de Valencia)

**Location:** Palau de la Música, Sala Iturbi. Avda. Alameda, 30. 46023 València

### » Wednesday 20<sup>th</sup> January, 16-18 h: Reception

Reception during the Research Speed Dating

### » Thursday 21<sup>th</sup> January, 21h: Training School dinner

**Restaurante Las Hermanas – Balneario La Alameda**

**Location:** C/ Amadeo de Saboya, 14. 46010 Valencia

### » Saturday 23<sup>th</sup> January: Trip to Albufera Natural Park

9.45 am Departure from *Hotel Puerta de Valencia* to Albufera

10.30 am Talk about the ecosystem of Albufera

Boat trip on the Albufera lake. The boat will stop by Trilladora del Tocaio.

1.30 pm Lunch at El Graner Restaurant

3.40 pm Visit to Trilladora del Tocaio

Projection of a documentary about "Trilladora del Tocaio"

5 pm Return to *Hotel Puerta de Valencia*

## Annex 5. “Discourse Relational Devices”, 3<sup>rd</sup> International Conference on Linguistic & Psycholinguistic Approaches to Text Structuring (LPTS 2016), Universitat de València, January 24-26, 2016.

### Sunday, 24 January

17:00 Welcome reception

18:00 Opening ceremony

18:15-19:30. Plenary: Bruce Fraser (Boston University) “Combinations of Discourse Markers”

### Monday, 25 January

8:30-9:00 Registration

	Room 106	Room 107 <i>Segmentation I</i>	Room 109 <i>Contrast I</i>
9:00-9:30	COLINET ET AL What does “Connective” Mean Cross-Linguistically?	SANDERS ET AL. Unifying Dimensions in Discourse Relations: how various annotation schemes are related	LEJEUNE The functional equivalents of French <i>Pourtant</i> in English and Portuguese
9:30-10:00	GABARRÓ When “same” means something different: the sign SAME as a discourse relational device in two sign languages	CRIBLE Position and context of structuring discourse markers: a corpus-based investigation of scope and cognitive load in French and English	VALŪNAITE-OLESKEVICIENE et al Adversative Conjunctions in Lithuanian and English
10:00-10:30	Poster session		
	BOLLY ET AL.: MDMA: corpus-based parameters for the identification of discourse markers in spoken French BORREGUERO, FIORENTINI, MARTÍNEZ CARO: Discourse markers in reported speech DOBROVOLJC Annotation of multi-word discourse markers in spoken Slovene DUPONT: Cohesive markers of contrast in English and French: what, where and why? JINOVA: Morphological and syntactic characteristics of discourse connectives in Czech NAGY: Rhetorical relations marked by ‘You know’ in the context of humor SÁNCHEZ CÁRDENAS: When Semantics and Pragmatics diverge: teaching students to translate adverbial connectors SANTANA COVARRUBIAS: The use of a Spanish Lexicon of Subjectivity for the analysis of Discourse Relational Devices (DRDs) in Spanish academic texts SCHOLMAN ET AL.: “On the one hand” as a Cue to Anticipate Upcoming DiscourseStructure		

10:30-11:00 Coffee break



11:00-12:30 Plenary: Liesbeth Degand (Université catholique de Louvain) "Discourse Relational Devices in TextLink: From (categorical) description to corpus annotation, and back again"

	Room 106 <i>Modality I</i>	Room 107 <i>Segmentation II</i>	Room 109 <i>Contrast II / Argumentative DRDs</i>
12:30-13:00	SAINZ Las conjunciones ilativas <i>así que</i> y <i>conque</i> como marcadores de modalidad evidencial	HOEK ET AL Segmenting discourse units: Incorporating interpretation into decision rules?	MATOS AMARAL & RODRÍGUEZ ROSIQUE Presupposition and semantic change: the case of 'pero'
13:00-13:30	GRZECH The Tena Kichwa =mi as a marker of speaker authority	SALAMEH Discourse markers in the Val.Es.Co. segmentation model: functions and uses of "pues eso" in Spanish colloquial conversations	SZCZYRBAK Concessive marking revisited: 'but' in courtroom talk
13:30-14:00	WEI & EVERS-VERMEUL Perspective marking and subjectivity in coherence relations: A collocation analysis of Chinese connectives	PONS BORDERÍA The combination of discourse markers and discourse units: not a morganatic marriage	VISCONTI Present Day Italian <i>anzi</i> in a cross-linguistic perspective

14:00-15:30 Lunch

15:30-17:00 Plenary: Óscar Loureda (University of Heidelberg) "Discourse markers and Experimental Linguistics"

17:00-17:30 Coffee break

	Room 106 <i>Modality II</i>	Room 107	Room 109
17:30-18:00	ALONSO ALMEIDA & CARRIÓN Modal and evidential devices as structure-signalling devices in early English texts	ATALLAH ET AL. Discourse relations, discourse connectives and discourse segmentation interdependency in the light of causality	CANES & DELBECQUE 'En realidad' as a crossing gate: multifunctionality and polysemy of a DM
18:00-18:30	LÓPEZ SERENA Las hipótesis sobre la gramaticalización de por cierto como MD epistémico en español: oralidad, escrituralidad y perspectiva cognitivo-funcional	HOFMOCKEL, FETZER & MAIER Discourse relations and strategic use of extra-clausal constituents in co-constructed argumentative discourse	LLOPIS What does <i>context</i> mean when referring to discourse markers?

## Tuesday, 26 January

	Room 106 <i>Phatic / interactive markers I</i>	Room 107 <i>Psycholinguistic approaches</i>	Room 109 <i>L2</i>
9:00-9:30	UCLÉS Phatic markers as conversation regulators in interviews and informal conversation	CRUZ RUBIO Processing of expected/unexpected focus in pragmatic scales in Spanish: experimental notes about the Spanish focus particles incluso	HOLGADO LAGE The different systems of organization of Spanish discourse markers and a new proposal for Spanish as an L/2
9:30-10:00	PASCUAL ALIAGA Hesitations, repetitions, restarts and self-corrections: Towards a systematization of discourse planning phenomena in Spanish colloquial conversations	ESPINOSA & GARCÍA Applying visualisation tools to analyse turn-taking systems in informal conversations and political interviews	LAITINEN Cognitive complexity in causal relations in adult L2-writing
10:00-10:30	SOUSA Contributos para a análise do conector "sim" em Português Europeu Contemporâneo	NADAL&RECIO Processing causality in Spanish: the case of "por tanto" and "por eso"	ELIODORO El marcador discursivo coloquial <i>bueno</i> : funciones pragmáticas en conversación coloquial descripción en manuales de E/L2 de nivel B2.

### 10:30: 11:00 Coffee break

11:00-12:30 Plenary: Volker Gast (University of Jena) "Doing corpus-based typology: Concessivity in a cross-linguistic perspective"

	Room 106	Room 107 <i>Contrastive Analysis</i>	Room 109 <i>Causality</i>
12:30-13:00	PADILLA HERRADA Marcadores interactivos en Twitter	LEWANDOWSKA & WILSON Negative Emotion Pragmatic Markers – Corpora and Cross-Linguistic Contrasts	FURKÓ The presence and absence of DRDs in Bible translations: the implicit and explicit marking of addition, contrast, and cause
13:00-13:30	PÉREZ ÁLVAREZ From Anaphora to Discourse Relational Devices	DE CESARE Les marqueurs discursifs de la relation d'ajout. Une étude contrastive français-italien	COKAL ET AL Causal Categories in Turkish connectives; First Results From a Corpus Study and a Judgement Task
13:30-14:00	POLÁKOVÁ Connective movement (A demonstration on Czech)	ALOTAIBI Comparison of Metadiscourse Markers in Arabic and English Research Articles in the	ARROYO HERNÁNDEZ Construcciones causales implícitas y explícitas en español e italiano

		Introduction and Conclusions Sections	
--	--	--	--

14:00-16:00 Lunch

16:00-17:30 Plenary: Kerstin Fischer (University of Southern Denmark)  
"Definitions of discourse markers and their functions as Discourse Relational  
Devices"

## Annex 6. Application form

Name:

E-mail address:

Contact phone number:

Affiliation:

Institution:

Department/School/Faculty:

Research Group:

Are you a:

- PhD student**
- Postdoc**
- Master student**
- Other (please specify):**

Are you a member of TextLink?

Supervisor, Principal Investigator of your group or other research leader:

Name:

E-mail:

Is he or she is a member of TextLink?

PhD or research topic area:

General information:

Do you have any experience on annotating:

- Discourse?**
- Discourse markers or other discourse relational devices?**

Which language(s) do you research?

Will you apply for a grant?

Note that only 30 grants will be awarded. If you are not awarded a grant, will you still be able to attend the school?

Briefly describe why you are interested in attending the TextLink Training School (maximum 200 words):



## Annex 7. Evaluation form

1st TextLink Training School

“Methods and Tools for the analysis of discourse relational devices”

(COST Action IS1312)

Universitat de València (Valencia, Spain), 18-22 January, 2016

<b>1. General comments about the 1<sup>st</sup> TextLink TS</b>
<b>2. Strong points</b>
<b>3. Aspects to be improved</b>
<b>4. Topics that you consider that should be included in the 2<sup>nd</sup> TextLink TS</b>
<b>5. Other aspects to be taken into account for the 2<sup>nd</sup> TextLink TS</b>

<b>Overall assessment of the 1<sup>st</sup> TextLink TS</b>			
Excellent	Very good	Good	Insufficient