UNIVERSITÀ DEL SALENTO

Dipartimento di STUDI UMANISTICI

To whom it may concern

We hereby declare that the paper "The development of DOTI (Data of Oral Teletandem Interactions). In D. Fišer & M. Beißwenger (Eds), *Investigating Computer-Mediated Communication: Corpus-based Approaches to Language in the Digital World* (pp. 172–192) is the result of a writing partnership. However, to accomplish some countries' academic requirements, we specify that Solange Aranha is responsible for sections 1., 1.1, 2., 2.1.1, 3., 3.2., 4, 4.2 and the abstract; Paola Leone is responsible for sections 1.2, 1.3, 2.1., 2.1.2, 3.1., 3.3, 4.1, and 5 of the current paper.

Solange Aranha

Associate Professor

Department of Modern Languages

Sao Paolo State University

Paola Leone

Paola Leone
Associate Professor
Department Human Studies
University of Salento

TRANSLATION STUDIES
AND APPLIED
LINGUISTICS

Edited by **Darja Fišer** and **Michael Beißwenger**

INVESTIGATING COMPUTER-MEDIATED COMMUNICATION:

CORPUS-BASED APPROACHES TO LANGUAGE IN THE DIGITAL WORLD

Edited by Darja Fišer and Michael Beißwenger

INVESTIGATING COMPUTER-MEDIATED COMMUNICATION: CORPUS-BASED APPROACHES TO LANGUAGE IN THE DIGITAL WORLD

Book series Translation Studies and Applied Linguistics

Ljubljana 2017

INVESTIGATING COMPUTER-MEDIATED COMMUNICATION: CORPUS-BASED APPROACHES TO LANGUAGE IN THE DIGITAL WORLD
BOOK SERIES TRANSLATION STUDIES AND APPLIED LINGUISTICS
ISSN 2335-335X

Edited by: Darja Fišer and Michael Beißwenger
Reviewers: Martina Ožbot, Irena Stramljič Breznik
Editorial board: Špela Vintar, Vojko Gorjanc and Nike Kocijančič Pokorn
English language proofreading: Paul Steed
Layout: Jure Preglau

Published by: Ljubljana University Press, Faculty of Arts
Issued by: Department of Translation Studies
For the publisher: Branka Kalenić Ramšak, Dean of the Faculty of Arts, University of Ljubljana

Ljubljana, 2017
Digital edition.
Design: Kofein, d. o. o.
Publication is free of charge.
DOI: 10.4312/9789612379612

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. J6-6842).

The book is avaliable in e-form (PDF) at https://e-knjige.ff.uni-lj.si/

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=292036352
ISBN 978-961-237-961-2 (pdf)

Table of contents

The development of DOTI (Data of oral teletandem interaction)

Solange Aranha, *Sao Paolo State University*
Paola Leone, *University of Salento*

Abstract

Teletandem[1] (Telles and Vassallo, 2006) is a Voice Over Internet Protocol (VoIP) communicative activity in which two speakers are involved, each of whom is an expert in one language and who wishes to learn the language of the interlocutor. "Virtual meetings" which last one hour are organized weekly; students speak half of the time in their L1, and the other half in the L2. Teletandem is also a growing field of research, and the related data, collected by video-recording the conversations between two participants, are an interesting resource for analysing communication and learning processes. In order to build a teletandem databank (DOTI – Data of Oral Teletandem Interactions), we collected data from Sao Paolo State University at São José do Rio Preto (Brazil: languages Portuguese/English), and from the University of Salento (Italy: languages Italian/English). DOTI is currently composed of about 700 hours of video data from, oral teletandem sessions. The current paper describes: i) the state of the art with regard to developing a databank with video recorded oral sessions, as well as chat conversations; ii) teletandem as an interaction space; iii) different learning scenarios and microtasks that might influence the type of data and, in turn, metadata, in this context.

Keywords: computer mediated interaction, databank, learner corpus, foreign language, learning scenario

1 Teletandem will be used with capital letters when we refer to the project "Teletandem Brasil – Foreign Language for all" and in lower case letters when we refer to practice/context/session.

1 INTRODUCTION

An interesting field of research in Applied Linguistics is the analysis of the various contexts in which L2 learning occurs, and the impact of all related variables on the development of L2 competence. The use of Information and Communication Technology (ICT) has recently created new opportunities for language learning worldwide, and many telecollaborative projects within universities have emerged in academic areas. Teletandem (Vassallo and Telles 2006), a telecollaborative project on language learning at the university level, is based on a multimodal form of interaction, carried out by the use of Voice Over Internet Protocol (VoIP) and Internet Relay Chat tool, aimed at promoting students' reciprocal learning. Two participants enrol in the activity and speak his/her language of proficiency for half of the oral session period, and for the other half the language he/she is learning. Such practice is based upon tandem principles, proposed by Brammerts in the 1980's (cf. Brammerts 1996), namely autonomy, separation of language and reciprocity. Autonomy implies the possibility that each participant has of organizing his/her own learning experience. Separation of languages means that only one language can be used for the part of the session dedicated to that language.[2] Reciprocity involves respect for the other's learning needs and commitment to practice both languages.

So far, Teletandem has led to many telecollaborative projects within universities, promoting networking for research purposes and the exchange of experiences and best practices. Nowadays, teletandem practice, as proposed by Telles and Vassallo (2006), is carried out in many universities around the globe: (seven in Europe: e.g. the University of Roma Tre, and Southampton University; twelve in the USA: e.g. Georgetown University, the University of Georgia in Athens, and Miami University, and two in South America: the University of Mexico and Cali University).[3]

The significance of such practice for language learning is twofold: it enriches the interactional skills of the participants through incidental learning, and grants them the possibility of sharing meaningful experiences in a dialogical and narrative path, which makes room for emphasizing relevant cultural characteristics. The teletandem experience allows participants to advance their linguistic-communicative competence as well as to expand their curiosity, to promote new themes, to question prejudices, to jeopardize discourses, and to discuss the interactional styles that characterize their cultures. Furthermore, it somehow establishes what Linnell (2009) calls the "sociocultural ecology" of linguistic learning, because the values and specificities of various cultures are not mediated by pedagogical materials and techniques, as is the case in a traditional language class (Telles, Zakir and Funo 2015). Because of this trait, teletandem implies a new type of mobility, achievable thanks to the use

2 Code-switching is, however, possible when it is aimed to facilitate conversations and messages (Leone 2009).

3 Information gathered among participating universities and their partner institutions (see www.teletandembrasil.org).

of ICT, i.e. *virtual mobility*, which works as a new way of "migration," even if temporary, to another country (Leone 2016). On the academic level, virtual mobility also supports future exchange programs (e.g. ERASMUS+).

The positive impact of teletandem on language learning is a good basis for expecting a gradual, but sustained increase in its use in higher education. This trend calls for further empirical research and implies a high demand for video/audio data.

Teletandem data, collected and filed using standard protocols which allow for systematic research, is required by users of the Teletandem network. In order to achieve this, we developed an arrangement between UNESP/SJRP and Unisalento with the ambition of filing and organizing existing data (Italian/English and Portuguese/English) in a databank composed of chat texts and video-recorded oral teletandem sessions, named DOTI (Databank of Oral Teletandem Interactions) (Aranha and Leone 2016).

The current paper is organized as follows. Section 1 relates teletandem to other CMC genres, reviews the literature on metadata in CMC corpora, and ends with the research questions. Section 2 describes the research context and DOTI project. Section 3 illustrates the main concepts used when defining the metadata for L2 interactions in pedagogical contexts (e.g. interaction space, learning scenario). Section 4 describes how those concepts are combined into DOTI metadata. Section 5 then concludes the paper.

1.1 Teletandem in relation to other CMC genres

Communication is generally synchronous during teletandem sessions, and quasi-synchronous when chat is employed. The typology of teletandem communication is defined both as telecollaboration and online intercultural exchange, according to Lewis and O'Dowd (2016), who intertwine the terms into a single meaning.

Teletandem practice implies multimodal spoken communication, and thus the data are both visual and vocal.[4] It provides a context for autonomous language learning, and is employed in institutions and sometimes even integrated into language courses.

Since during teletandem sessions the participants talk while keeping in mind a double focus, the language used and the discussion themes (Apfelbaum 1993; Bange 1992; Leone 2014a), teletandem is defined as "conversation for learning" (Kasper 2004; Kasper and Younhee 2015). As a pedagogical context, when collecting data for research purposes, the features of the learning situation need to be described

4 This type of data generates problems with regard to privacy, which have been dealt with by asking the participants to sign a consent form.

and the characteristics of groups and participants must be recorded. For example, we must take into account the organization of a teletandem activity (e.g. length of the program), the learning situation (e.g. the presence or absence of a task), students' sociodemographic profiles (e.g. gender, age), because research shows that these properties might affect how participants interact (see Rampazzo 2017).

The main features of teletandem sessions (i.e. being spoken multimodal communication carried out in a learning institution) differentiate this form of telecollaboration from other CMC exchanges, such as conferencing systems communication (conference systems with text, etc.), email, discussion forums, blogs, tweets, and audio-graphic systems. These are written exchanges and they are not seen as empowering the users' language and cultural skills and, in most cases, they are not so strongly linked with a learning institution.

To the best of our knowledge, few multimodal data have been used for databank and corpora building. For instance, Chanier et al. (2014) describe the Corpus de Communication Médiée par de Réseaux (COMERE)[5] which covers different genres. Chanier and Wigham (2016) describe procedures used with the Learning and Teaching Corpora (LETEC), based on previous experiences with the Mulce projetc. In both cases, learning environments are currently scheduled to be included in the related corpus. Other corpora described in the recently published volume Wigham and Ledegen (2017) do not include either data from computer mediated learning contexts, nor spoken interaction data. In the pedagogical domain, Mangenot and Soubrié (2010) discuss the development of a learning objects' databank as an open resource, highlighting the importance of "task" as a unit for describing teaching practices. The shortage of such data is probably due to the fact that such learning experiences are recent, and the transcription procedure is still time consuming, even if transcription software (e.g. Transana, ELAN) now supports this. Nevertheless, according to Chanier and Wigham (2016: 216):

> Studying online learning, in order to understand this specific type of situated human learning (Learner Computer Interactions (LCI)) and/or evaluate pedagogical scenarios or technological environments, requires accessibility to interaction data collected from the learning situation.

1.2 CMC and metadata

Metadata are "management tools" (Autayeu, Giunchiglia and Andrews 2010) which allow users to process and select relevant data. For browsing the web and looking for a journal article, for instance, we can write two or more words of the paper's title.

5 See https://corpuscomere.wordpress.com/

The titles of papers and books, keywords or business catalogues' names are manually generated natural language metadata. Conversely, the date of a picture is automatically generated by the camera. Natural language or standardized metadata are listed in different datasets, each including "web directory category names, business catalogue category names, thesauri and subject headings" (Autayeu et al. 2010). Datasets can be more general or specific to a certain domain. For example, DMoz or Open Directory Project is quite general, very large and used as a directory for classifying all sites, including well-known search engines such as Google. The Dublin Core is "a vocabulary of fifteen properties for use in resource description".[6] On the other hand, the Text Encoding Initiative (TEI; Burnard and Bauman 2013) encodes metadata for machine-readable texts and is used in the field of humanities, social sciences and linguistics, while LOM (Learning Object Metadata) and SCORM (Sharable Content Object Reference Metadata) are applied in that of pedagogy.

Metadata are characterized by "atomic concepts," with Autayeu et al. (2010) noting that the query "Bank and personal detail of George Bush" is made of four atomic concepts: bank, personal, detail and "George Bush". "Atomic concepts" are thus used to create complex concepts (Autayeu et al. 2010).

Most standardized metadata need to be extended in order to encompass more recent computer mediated texts and learning experiences. For this reason, careful and focused illustration of different computer mediated learning environments and practices is needed to create a model stemming from the highlighted characteristics (Mangenot and Soubrié 2010).

1.3 Research questions

Much research has been carried out within the teletandem learning context (cf. www.teletandembrasil.org). The list of published works in this area emphasizes the coverage of multiple theoretical perspectives and presents a fertile field for understanding how telecollaboration may enhance participants' competences. If the wide inventory of pedagogical experiences and scientific studies has enriched the original project with new interpretations and perspectives, one current task is to better understand how the initial proposal by Telles and Vassallo (2006) has been actually carried out in various international contexts.

The present research is a first step in this direction and tries to meet the urgent need to reflect upon what has been done, starting from two academic contexts: UNESP (Sao Paolo State University) and Unisalento (Universidad del Salento). Such a simplified but comprehensive description is also used to describe the

6 See http://dublincore.org/

amount of data which has already been generated and recorded, and which can be further collected and filed within the project.

Bearing the above in mind, the current study aims at answering the following questions:

1) How can teletandem exchanges be encoded in standardized metadata?
2) What are the common characteristics of the learning contexts as they are developed in the Brazilian and the Italian higher educational institutions examined in this work?
3) Which metadata allow the identification of online interactions with learning purposes?

We intend to follow a common course to establish metadata for describing DOTI, as well as take a first step towards the definition of a protocol for collecting further data, and transcribing existing data.[7] This work has two main aims: a) to enhance collaborative and shared research among Teletandem network members; and b) to expand and reinforce the network between professors and mediators.

2 RESEARCH CONTEXT

Teletandem practice may occur within a language course, as part of a university program, as seen in some groups at Sao Paolo State University at São José do Rio Preto (UNESP/SJRP), or may occur as an elective activity, thus voluntary, as in the University of Salento. The former is coined institutional integrated teletandem, and the latter institutional non-integrated teletandem, according to Aranha and Cavalari (2014). Depending on the agreement between the two partner institutions that carry out a teletandem program, the computer mediated oral sessions may or may not be followed by other learning activities or tasks.

In our universities, teletandem practice has been adopted with students from different majors who study various foreign languages. Their level of L2 linguistic competence varies, although this is not taken into account if a person wants to join the Teletandem project. The tasks that can occur in diverse learning scenarios are adjustable to distinct levels of competence. The characteristics of a teletandem course can be described following: a) a general framework for identifying the communication setting (Mangenot and Soubrié 2010); b) a general framework for outlining both pedagogical and learning practices based upon teletandem at UNESP-SJRP and at Unisalento.

7 10% of existing data (Portuguese/English) has already been transcribed by using Transana.

2.1 UNESP and Unisalento specific characteristics of Teletandem Oral Sessions and Mediation

Teletandem practice displays a learning scenario that carries a coherent and complex activity framework – a TOS (Teletandem oral session) and teletandem mediation session – which consists of different pedagogical and didactic collaborative events (Mangenot 2008, Foucher 2010) aimed at developing students' plurilingual and pluricultural competences (Candelier et al., 2012, Leone et al. 2015).

Teletandem pedagogical scenarios (TTPS) are coherent with the following principles: (i) collaboration: the tasks are intended to be developed collaboratively; ii) interaction: communicative exchanges and oral sessions favour the development of learning strategies and autonomy, and also increase inter-comprehension skills.

TTPSs have varied purposes, which can be synthetized into four points. The first has the intent of preparing students to participate actively in (computer mediated) oral interactions with a proficient speaker, and be aware of all the linguistic and cultural strategies that such a practice involves (Aranha and Leone 2016). The second aims at improving self-evaluation and awareness about one's own learning skills and abilities.

The objectives are, therefore, to make the participants more autonomous in their learning and then to develop their "learning how to learn skill." Or better yet, to be aware of how to study and improve/articulate knowledge and competences outside of formal contexts and without teaching guidance. The third point, the scenarios based on teletandem, has the purpose of promoting the use of digital technology to facilitate one's learning capacity efficiently and flexibly. In this sense, the participants may take advantage of the great potential of new technologies, considered as key-knowledge tools in the Recommendation of the European Parliament and of the Council of 2006.[8] Finally, through intercultural discourse, TTPSs give the participants opportunities to strengthen their positive attitudes towards other people, ideas, experiences and cultures.

Teletandem is characterized by two macrotasks: mediation sessions and teletandem oral sessions (TOS). At UNESP/SJRP, mediators are both professors involved in the Teletandem project and graduate students (Masters and Doctorates) who investigate telecollaboration practices. At Unisalento, mediators are language instructors and professors involved in the project.

The linguistic and cultural exchanges between mediators and students happen within a social cultural perspective and allow each and every individual to advocate cultural identities in a broad sense. During mediation sessions, participants interact and "do

8 Recommendation of the European Parliament and of the Council of 18th December 2006 on key competences for lifelong learning http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32006H0962

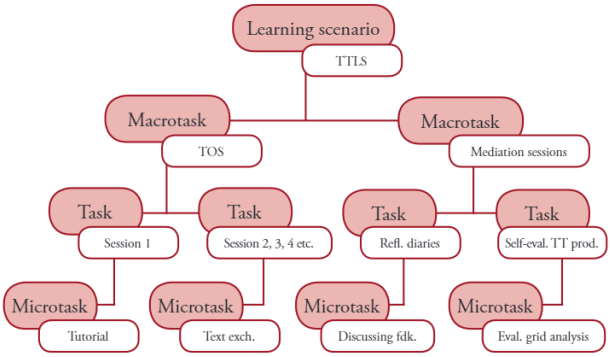

Figure 1: The organization of a pedagogical scenario based on Teletandem.

not simply shift between competing meanings to find the correct one, but, instead, navigate a constantly changing and emerging hermeneutic environment." (Feito 2007: 3). Teletandem principles and environments allow students to negotiate meaning, discuss points of view, envision new knowledge, and present cultural approaches and perspectives.

The length of learning scenarios based on the teletandem context is variable (from six to 15 sessions, of about one hour each, depending on the needs of each group). Online meetings are video recorded using Evaer[9] and consent forms are signed to protect the privacy of participants and allow the research process to be controlled (Aranha et al. 2015, Mackey and Gass 2005: 330). As shown in Fig. 1, for describing our learning scenarios we use the terms *tasks*, *macrotasks* and *microtasks* to present the complex environment in which teletandem practice takes place. As argued by Mangenot and Soubrié (2010), the concept of task is essential for developing descriptors of more recent teaching practice. *Macrotasks* are tasks with a larger scale and scope, involving teletandem sessions and mediation sessions. *Microtasks* are short duration tasks with reduced scopes, and these support overall task implementation.

In TTPS, technology is used both to carry out oral exchanges (via VOIP technology) and to develop access to documents and assigned activities within each scenario (e.g. Moodle, Google Docs). Technologies are also essential for some *microtasks*, such as collaborative writing, or recording and analysing recorded videos. TTPS offers, therefore, the use of technology for communication (Computer Mediated

9 Evaer is an easy-to-use and low-cost software package for recording Skype calls. See http://www.evaer.com/

Communication: CMC) and for assisting students in their learning process (Computer Assisted Language Learning: CALL).

2.1.1 Pedagogical scenarios at UNESP/SJRP

At UNESP/SJRP, students who participate in the Teletandem Based Learning Scenario (TTPS) are from different linguistic levels and majors. In the integrated scenarios, students are majoring in language courses and aiming to be either be language teachers or translators. In non-integrated ones, undergraduates come from different courses, and their alleged level of proficiency is self-established, i.e. using the grid from the Common European Framework (CEFR), students place themselves in one level when they answer a questionnaire before the sessions begin. TOSs are fed by texts exchanged between partners and guided by some pedagogical tasks. Free conversation also occur.

2.1.2 Learning scenario at Unisalento

TTPS are institutionalized but non-integrated at Unisalento, and credits are awarded for participation and completion of tasks. Students who participate in learning scenarios based on teletandem attend Bachelor's and Master's degree courses, and specialize in one or more foreign languages (e.g. English, French, and Arabic). As at UNESP, language competence is self-established by students using CEFR grids for evaluation.

TOSs are (currently) characterized by free conversations or by discussions on specific topics (e.g. youth life-styles in the students' countries; Leone 2016).

3 METHODOLOGY

Teletandem practices in the two higher educational contexts had to be shared so that we defined the pedagogical characteristics of such learning practices, trying to uncover those which could allow the description of the whole process. For describing Teletandem sessions, the notion of "interaction space", as developed in Chanier et al. (2014), was used. Since teletandem is a pedagogic and communicative practice in which students and professors are both involved, the concept of learning scenario (Mangenot 2008, Foucher 2010) must also be present as an epistemological frame, useful for characterizing various sequences and events that determine it.

The components of different learning scenarios (e.g. characteristics of participants, number of sessions), microtasks (e.g. methodological procedures, verbal and non-verbal input), as well as the properties of the interaction space, within which the various forms of technology mediated communication are performed for completing the learning scenario, are all considered. Concerning the pedagogical implementation of Teletandem, we developed a didactic description that is a first step in the process of producing standardized metadata.

In the following subsections we examine the concepts of interaction space, learning scenario and task in more detail.

3.1 Interaction space

The notion of Interaction Space (IS) (Chanier et al. 2014) derives from TEI and aims at characterizing distinct genres within CMC (focused on written communication, such as Facebook posts), and is defined as an abstract concept "located in **time [...]** where interactions between **a set of participants** occur within an **online location**".

As described by Chanier et al. (2014), IS entails concepts related to *Interaction Space* itself and to *CMC environment*. The first includes *participants*, i.e. a set of groups or individuals, *time frame*, i.e. the beginning and ending time, and *online location*. *CMC environment* gives access to online communication, and it can be *monomodal* or *multimodal*. *Modality* is "a specific way for realizing communication" (Chanier et al. 2014: 6), and it affords a specific *interaction type* (e.g., email). Modality can also be described in terms of "semiotic resource", that is the *mode* (i.e., text, speech and non-verbal) which realizes communication. Finally, *time* can be synchronous or asynchronous.

3.2 Pedagogical and learning scenario

For describing online learning situations, Chanier and Wigham (2016: 222) use the term pedagogical scenario. A pedagogical scenario describes:

a) the whole environment (such as a Learning Management System (LMS);
b) the various roles of participants (teachers, learners, experts and the role of each participant during the course);
c) each course activity and the role of each participant during this;
d) how activities are sequenced;
e) the resources that will be used and produced; and
f) the instructions that govern the learning activities.

A pedagogical scenario may consist of a learning scenario and a tutoring/supervision scenario. Using Chanier and Wigham's terminology, DOTI is, so far, composed only of one learning scenario (which we call macrotask), although the Telétandem project also considers a tutoring/supervision scenario, as described below.

3.3 Task as an essential concept

For us, a task is an essential unit for defining all the activities carried out in a learning scenario, and thus a task can be considered one of our "atomic concepts" (see par. 1.2). As above mentioned, the concept of a task is also essential to describe specific activities, such as microtasks.

Many definitions of a task appear in the literature, and all of them imply that any effective task integrated in formal educational programmes must be communicative, meaning-focused and linked to the real (i.e., beyond the classroom) use of that language (Skehan 1998). According to Gonzales-Llore and Ortega (2015), the primary focus of a task is on meaning. Even if there is a preplanned language learning goal, part of the learning must be incidental, and any particular language focus should be hidden from the learners, or 'implicit,' at least for a good part of the task module. Long (2015: 3470) emphasizes that classroom tasks[10] should be based on students' learning needs, definable by the activity they "need, or will need, to do in the L2," which Long terms "target task" (Long 2015: 3479). Gonzales-Llore and Ortega (2014: 5) mention holism as one definitional feature of a task in the context of technology-and-task integration:

> a task draws on real-world processes of language use, integrating form-function-meaning; this definitional feature goes to notions of 'authenticity' and 'real-world relationship'.

In our experience, we believe that autonomy in L2 learning is crucial for our students' future professions, and thus the main task of mediation sessions is based on a target task, which is "self-evaluating one's interaction skills and analysing the learning process." In fact, we aim at developing students' abilities to self-analyse their own learning process and the communicative use of "the lexis, collocations, pragmatics, skills, genre and registers" (Long 2015: 3466) necessary for reflecting on their own L2 production, learning process and needs.

Following the framework by Ellis (2003) and Gonzales-Llore and Ortega (2015), the two tasks of the learning scenarios in the two higher educational contexts examined in this work, i.e. diaries and self-evaluating interaction skills, can be described based on the following design features (see also Mangenot and Soubrié 2010):

10 Although Long consider "classroom" tasks and our context is not within a classroom, we argue that the concept also applies to telecollaborative practices.

1. Goal (intended as the general purpose of the task). The task plan must offer a language-and-action experience, which means the task must entail (a) some communicative purpose (i.e. considering students' needs and wants) engineered by means of gap in information or some element that encourages language use that involves informational transfer; and (b) some outcome, resulting from task completion, including communicative outcomes (e. g. the production of an oral or written message, the accomplishment of a desired perlocutionary effect on interlocutors or on the world) and /or non-communicative outcomes (securing a flight booking, producing a plan, gathering knowledge, playing/winning a game, and so on). The goal is the development of autonomy in L2 learning.
2. Input, which may mean the verbal and non-verbal information provided for the task: websites, tutorials, previous learning experience, epiphanies, diaries, teletandem session video-recordings and the CEFR evaluation grids.
3. Conditions are how the information is provided. Normally students do not share the same information. For instance, in "self-evaluating interaction skills," each student does not know which video sequence his/her partners are going to show and comment on. In diaries, one-to-one feedback is given by the professor or mediator in charge and, although much information from the diaries is used for the group mediation meeting, much is personal and directed to one individual.
4. Procedure (e.g. group work vs. pair work; planning time vs. no planning time), at both UNESP/SJRP and Unisalento students work individually during the TOS learning scenario. Afterwards, at UNESP, they share their views in the reflexive diaries, which may be used by professors for classroom and mediation purposes. At Unisalento each student self-evaluates their production and discusses it with the mediator, as well as with thers.
5. Outcomes. Diaries are the products at UNESP/SJRP, while at Unisalento the focus is on an oral discussion of the experience supported by a presentation file. For both tasks, the process of the linguistic interaction and the cognitive activity generated by the task have a strong educational value.

4 DOTI characteristics and metadata

DOTI is composed of around 700 hours of teletandem oral sessions,[11] one of the learning scenarios described above. The majority of these sessions were collect-

11 At UNESP, the texts produced within the macrotasks, tutorials, questionnaires and reflexive diaries are part of another databank.

ed from the Brazilian university, carried out in Portuguese/English. Unisalento provided fewer recorded oral sessions, with the TOSs being in Italian/English. However, the fact that this data is unbalanced in terms of number of hours for each pair of languages should not be seen as a weakness, as DOTI is ultimately intended as a multilingual databank.

Due to the attributes of DOTI, the databank will provide input to answer the following types of research questions: What are the differences between chat and video synchronous communication? What are the aspects that distinguish chat and oral communication in a learning environment and other contexts (e.g. among friends)? What are the distinctive features of teletandem oral session in relation to other types of oral communication between native and non-native speakers? What are the typical features of metalinguistic sequences in teletandem oral sessions and other virtual contexts (e.g. forums)? Which are the genres used for teletandem interactions in various modalities and microtasks? Which genres are typical of telecollaborative practice? Are genres related to learning scenarios?[12] How do the genres that occur within a teletandem context relate to cultural and linguistic learning?

In sections 4.1 and 4.2 we present metadata concerning the interaction space and the learning scenario. The former shows the general characteristics of teletandem oral sessions, and the latter presents a rough outline of pedagogical issues related to the formative path based on Teletandem.

4.1 Teletandem as an interaction space

In relation to the interaction space, Teletandem is characterized in terms of *participants*, *place/institution* and *time frame*. The participants of TOSs will be two students who want to learn the language of his/her partner; the institutions may be UNESP and UGA (University of Georgia), or UNISALENTO and other British or American universities; *place/institution* records the names of the institutions involved; *time frame* will include information on the semester/year, number of sessions and duration of each session.

In relation to *technology environment*, teletandem is multimodal (visual, oral and written), synchronous as opposed to asynchronous online communication (e.g. blogs). Moreover, the *language* used (e.g. English and Italian, Portuguese and English) will also be specified.

12 Rampazzo's thesis (2016) shows that the Initial Teletandem Oral Session, as a genre, is dependent on the related learning scenarios.

4.2 Pedagogical scenario

The descriptors will be: *pedagogical scenario*, *macrotasks* (i.e. TOS and mediation sessions), *task* (e.g. learning diaries) and *microtask*, and thus the metadata subfields for the two universities examined here will be different. For example, at UNESP/SJRP teletandem is integrated in a course syllabus, while at Unisalento it is not. Because of this, integrated or non-integrated modalities are also taken into account. If it is non-integrated, then any credits that are awarded should also be included in the data.

Concerning the learning scenario, all the information combined in the following fields and subfields are considered: the *university curriculum* – with an integrated or non-integrated modality; *time frame*, indicating when and for how long TTPS happened. *Pedagogical scenario* (Fig. 1) also entails the *number* of *macrotasks* and *typology* (e.g. teletandem sessions and mediation sessions). In the following section, we will focus on metadata concerning the Teletandem macrotasks, while mediation macrotasks are not considered since they are currently not part of DOTI.

Teletandem metadata clarifies characteristics related to the learning scenario and teletandem sessions. For the learning scenarios we created a template (Fig.2) that includes information on: 1) learning scenario modality (i.e. integrated, non-integrated); 2) institutions involved; 3) students' majors; 4) professors; 5) mediators; 6) periods of mediation; 7) length of teletandem activity; 9) number of interactions; and 10) place.

Teletandem sessions are described considering, first of all, the participants, based on their sociodemographic characteristics and university curricula. With regard to the CMC environment, Teletandem is multimodal. The mode, i.e., the semiotic source, is text (chat), speech and non-verbal. The interaction type is oral. Finally, time is synchronous (video-conference) and quasi-synchronous (chat).

In terms of TOS, we created another document that includes sociodemographic characteristics, including information about each participant, i.e. *Major*, *Gender* (F or M and Other), and *Alleged Language Competence level* in L2. The pedagogical characteristics of Teletandem are described in terms of *task* and *discourse type* (e.g. free conversation; discussions about a specific theme; development of a task).

The description of a task will include the goal, input, conditions, and the related procedures will also be described.

teletandem brasil
línguas estrangeiras para todos

MODALITY	INSTITUCIONAL INTEGRADO	
INSTITUTIONS	UNESP	SHEFFIELD
CLASSES		
PROFESSORS	SOLANGE	CARMEM
MEDIATORS	Fernanda	X
MEDIATION	Each 2 weeks	Not expected
PERIOD	March 24, 2017 to May 12, 2017	
DAY	FRIDAY	
TIME		
MARCH	9:00	12:00
APRIL/MAY	8:00	12:00
TOSs #	8	
PLACE	TTD Lab	Lab
DISCOURSE TYPE	Free conversation	Specific theme discussion
TYPOLOGY	Alternate monolingualism	

Observe: Discourse type: Free conversation/ Task realization/ Discussão specific theme

Figure 2: Document for describing a teletandem learning scenario at UNESP.

5 CONCLUSION

The study examined a specialized segment of computer mediated research, as collecting, organizing and sharing spoken oral data for language learning is an emerging field of research in CALL. More specifically, this study aimed to develop a databank, named DOTI, composed of approximately 700 hours of TOS, and presented several descriptors generated from two key concepts: Interaction Space and Learning Scenario. The former places DOTI within a broader context that includes resources and research on other forms of CMC (such as Facebook and Twitter). The latter is, instead, used to outline the distinctive features of the academic and educational contexts in which teletandem is practiced. When defining metadata, the concept of task, a unit for describing the learning scenario, proved to be significant. Moreover, the metadata used for the learning scenario need to be developed into more standardized forms.

Since every year new partnerships are formed, a growing body of experience can be used to define the agreements that occur between new partner institutions. This first step of this study at creating guidelines for developing the proposed databank will help other researchers to develop more reliable tools for future research. For this

reason, the proposed metadata will also be used to establish a protocol of collecting and filing new data. The protocol will be used to: a) save time in collecting data by members of the network; b) share collecting and transcription methodologies; c) enhance the use of sound, scientific procedures. Once the databank is transformed into a LETEC (Learning and Teaching Corpora) corpus, the data can then be interrogated by multiple researchers and for various purposes.[13]

References

Apfelbaum, Birgit, 1993. *Erzählen im Tandem. Sprachlernaktivitäten und die Konstructio eines Diskursmusters in der Fremdsprache (Zielsprachen: Französisch und Deutsch).* Tübingen: Narr.

Aranha, Solange and Spatti Cavalari, 2014. A trajetória do projeto Teletandem Brasil: da modalidade institucional não-integrada à institucional integrada. *The ESPecialist* 35/2, 70–88.

Aranha, Solange and Paola Leone, 2016. DOTI: Databank of Oral Teletandem Interactions. Jager, Sake and Malgorzata Kurek (eds.): *New directions in telecollaborative research and practice: selected papers from the second conference on telecollaboration in higher education*. 327–332.

Aranha, Solange, Lidiane Luvizari-Murad and Augusto César Moreno, 2015. A criação de um banco de dados para pesquisas sobre aprendizagem via teletandem institucional integrado (TTDII). *Revista (Con) Textos Linguísticos* 9/12. 274–293.

Autayeu, Aliaksandr, Fausto Giunchiglia and Pierre Andrews, 2010. Understanding Natural Language Metadata. http://eprints.biblio.unitn.it/1836/1/026.pdf. (Last accessed 29 June 2017.)

Bakhtin, Mikhail, 1986. *Speech genres and other late essays. A selection of essays from the Russian original "Estetika slovesnogo tvorchestva" [1979].* Austin: University of Texas Press.

Bange, Pierre, 1992. A propos de la communication et de l'apprendissage en L2, notamment dans le forme institutionnelles. *Aile* 1. 53–55.

Brammerts, Helmut E., 1996. Tandem language learning via the internet and the International E-Mail tandem network. Little, David and Helmut E. Brammerts (eds.) *A guide to language learning in tandem via the internet.* CLCS Occasional Paper no. 47. Dublin: Trinity College Dublin. 9–22.

Burnard, Lou and Syd Bauman, 2013. TEI P5: Guidelines for electronic text encoding and interchange. http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines. (Last accessed 29 June 2017.)

13 We would like to mention that one of authors has been awarded a grant (FAPESP #2016/18705-9) that will help fund the organization proposed for developing DOTI. Moreover, on the Brazilian side, the various other microtasks (diaries, texts exchanged between partners, questionnaires) are also part of the databank.

Candelier, Michel, Antoinette Camilleri-Grima, Véronique Castellotti, Jean-Francois de Pietro, Ildikó Lörincz, Franz-Joseph Meissner, Anna Schröder-Sura and Artur Noguerol, 2012. *CARAP. Un Cadre De Référence pour les Approaches Plurielles des langues et des Cultures. Compétences et ressources.* CELV (Centre Européen pour le Langue Vivantes). http://apfmalte.com/uploads/CARAP.pdf. (Last accessed 10 August 2017.)

Cavalari, Suzi M. and Solange Aranha, in preparation. *Implications of Teletandem integration into foreign language programs: insights on the teacher-mediator's role.*

Cavalari, Suzi M. and Solange Aranha, 2016. Teletandem: integrating e-learning into the foreign language classroom. *Acta Scientiarum: Language and Culture* 38/4. 327–336.

Chanier, Thierry, Céline Poudat, Benoit Sagot, Georges Antoniadis, Ciara R. Wigham, Linda Hriba, Julien Longhi, Djamé Seddah, 2014. The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres. *Journal for Language Technology and Computational Linguistics* 2/29. 1–30.

Chanier Thierry and Ciara Wigham, 2016. A scientific methodology for researching CALL interaction data: Multimodal LEarning and TEaching Corpora. Caws, Catherine and Marie-Joseé Hamel (eds.): *Language-Learner Computer Interactions: Theory, methodology and CALL applications*, Johns Benjamins.

Feito, José A., 2007. Allowing not-knowing in a dialogic discussion. *International Journal for the Scholarship of Teaching and Learning* 1/1. 1–11.

Foucher, Anne-Laure, 2010. *Didactique des Langues-Cultures et Tice : scénarios, taches, interactions.* Université Blaise Pascal - Clermont-Ferrand II.

Gonzales-Llore, Marta and Lourdes Ortega, 2015. *Technology-mediated TBLT. Researching Technology and Tasks.* Amsterdam/Philadelphia: John Benjamins.

Kasper, Gabriele, 2004. Participant orientations in German Conversation-for-Learning. *The Modern Language Journal* 88/4. 551–567.

Kasper, Gabriel and Kim Younhee, 2015. Conversation-for-Learning: Institutional Talk Beyond the Classroom. *The Handbook of Classroom Discourse and Interaction*. London: Wiley Blackwell. 390–408.

Leone, Paola, 2016. Collaborare per capirsi nel contesto di apprendimento teletandem. *Parlare insieme. Studi per Daniela Zorzi*. Bologna: Bononia University Press. 191–206

Leone, Paola, 2014a. Focus on form durante conversazioni esolingui via computer. *Varietà dei contesti di apprendimento linguistico*. Milano: Officinaventuno. 169–187

Leone, Paola, 2016. Migrazioni virtuali: teletandem per l'apprendimento di una L2. *Incontri*. 48–65.

Leone, Paola, 2009. Processi negoziali nel corso di scambi comunicativi mediati dal computer. *Oralità/scrittura. In memoria di Giorgio Raimondo Cardona*. Perugia: Guerra. 389–412

Leone, Paola and João Telles, 2016. The Teletandem network. *Online Intercultural Exchange: Policy, Pedagogy, Practice*. London: Routledge. 243–248.
Leone, Paola, Alessandro Bitonti, Donatella Resta and Bianca Sisinni, 2015. *Osservazione di classe, insegnamento linguistico e (tele)collaborazione.* Firenza: Franco Cesati.
Lewis, Tim and Robert O'Dowd (eds.), 2016. *Online Intercultural Exchange: Policy, Pedagogy, Practice.* London: Routledge.
Linell, Per, 2009. *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-making.* Charlotte, NC: Information Age Publishing.
Long, Michael, 2015. *Understanding second language acquisition.* Oxford: Oxford University Press.
Luvizari-Murad, Lidiane H., 2011. *Aprendizagem de alemão e português via teletandem: um estudo com base na Teoria da Atividade.* Unpublished PhD Thesis.
Mackey, Alison and Susan M. Gass, 2005. *Second language research: methodology and design.* Mahwah, NJ: Lawrence Erlbaum.
Mangenot, François, 2008. La question du scénario de communication dans les interactions pédagogiques en ligne. *Jocair (Journées Communication et Apprentissage Instrumentés en Réseau*. 13–26.
Mangenot, François and Thierry Soubrié, 2010. Créer une banque de tâches Internet: quels descripteurs pour quelles utilisations? *La tâche comme point focal de l'apprendissage.* Clermont-Ferrand.
O'Dowd, Robert, 2013. The competences of the telecollaborative teacher. *The Language Learning Journal*. 194–207.
Rampazzo, Laura, 2017. *Gêneros textuais e telecolaboração: uma investigação da sessão oral teletandem inicial*. Dissertação (Mestrado em Estudos Linguísticos). Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de São José do Rio Preto. São José do Rio Preto.
Skehan, Peter, 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.
Telles, João and Maria L. Vassallo, 2006. Foreign language learning in-tandem: Teletandem as an alternative proposal in CALLT. *The ESPecialist* 27/2. 189–212.
Telles, João A., Maisa de Alcântara Zakir and Ludmila B.A. Funo, 2015. Teletandem and culture-related episodes. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 31/2, 359–389.
Vygotsky, L.S, 1978. *Mind in Society.* Cambridge, MA: Harvard University Press.
Wigham, Ciara and Ledegen Gudrun (eds.), 2017. *Corpus de communication médiée e par les réseaux: construction, structuration, analyse*. L'Harmattan, Collection Humanités Numériques.