

LANGUAGE SEPARATION IN TELETANDEM: WHAT CORPUS ANALYSIS CAN TELL US

Rodrigo E. de Lima-Lopes^{1*}

¹Universidade Estadual de Campinas, Campinas, SP, Brasil

Solange Aranha^{2**}

²Universidade Estadual de São Paulo, São Paulo, SP, Brasil

Abstract

The purpose of this paper is to investigate 16 Initial Teletandem Oral Sessions (iTOS) within Teletandem Brasil Project to verify if the participants rely on language separation, one of the principles adopted by tandem learning. We focus on the lexical choices Brazilian students learning English as a Foreign Language, and US (non-Brazilian) citizens learning Portuguese as a Foreign Language make during their conversation. Data are from MulTeC (Multimodal Teletandem Corpus) and the lexical choices in each language segment were manually classified using wordlists and Keyword in Context and ANOVA tests were calculated across the sessions. Results show that the language separation principle tends to be respected and that Brazilians and non-Brazilians continue speaking the target language during the sessions; however, they also use a myriad of resources to maintain communication. This article contributes to the study of teletandem, primarily by quantitatively analysing the linguistic data of initial sessions thus making it possible to understand some of the participants' strategies to overcome their linguistic difficulties.

Keywords: Separation of languages principle, Corpus Linguistics, Learner Corpora, Initial Teletandem Oral Sessions (iTOS).

* Associate Professor of Applied Linguistics at the Institute of Language Studies at UNICAMP (State University of Campinas) and a CNPq (Brazilian National Council for Scientific and Technological Development) fellow supported researcher. He works on both Undergraduate and Graduate Programmes in Applied Linguistics and his research interests focus on Digital Humanities, Datification of Society, Discourse Analysis and Systemic-Functional Linguistics. Email: rll307@unicamp.br ORCID: <https://orcid.org/0000-0003-3681-1553>.

** Associate Professor of English at the Modern Languages Department at UNESP (Sao Paulo State University). She teaches both undergraduate and graduate levels and advises Masters and PhD students in the Applied Linguistics field. Her research focuses on genres, telecollaboration and teaching and learning foreign languages. Her research has been sponsored by FAPESP (Sao Paulo Research Foundation). Email: solange.aranha@unesp.br ORCID: <https://orcid.org/0000-0002-8092-1875>.



1. Introduction

The Teletandem Brasil Project - Foreign Languages for all (Telles, 2015) was designed to be implemented via CMC (Computer Mediated Communication) tools by following the tandem learning approach to learning foreign languages, which has been used in Europe since the 60's. In teletandem, two proficient partners meet regularly so that one can help the other learn his/her language. (Tele) Tandem learning is guided by three principles: autonomy, reciprocity, and separation of languages (Brammerts, 1996). The first one postulates that each participant should make decisions about their own learning process and be responsible for them. The second one is usually associated with the amount of interaction in each language and with the commitment to one's partner's learning preferences. This principle assures that both participants have equal and collaborative conditions of exchange, once both are given similar backgrounds to practise the language they are learning and "teach" their language of proficiency. This lack of hierarchy aims at improving participants' self-esteem and confidence to speak the target language. The third principle is related to the amount of time dedicated to each of the languages and to the indication that both languages should not be mixed, but solely used for the specified time, i.e., half of the teletandem oral session (TOS). This separation may assure that each participant looks for communication strategies to overcome linguistic barriers, which means that instead of using their mother tongue when one does not know a word or expression, one should stick to the target language to communicate. Vassallo and Telles (2006, p. 23) state that the objective of this principle is to "challenge students to speak the target language".

Research has focused on different aspects of the practice. Theses, dissertations and papers have brought aspects of learning a foreign language in telecollaboration into attention (cf. <http://www.teletandembrasil.org/>). One of the relevant aspects to be investigated within the teletandem context is the respect to tandem principles (reciprocity, autonomy and separation of languages) during the Teletandem Oral Session (TOS) and the interconnection of them (Cappellini et al., 2019; Elstermann, 2017; Picoli & Salomão, 2020, among others).¹ Although the three of them are intertwined in practice, separation of languages puts tandem as a bilingual learning environment (Lewis, 2020). This principle guarantees what Brammerts (2003) indicates as a crucial aspect of tandem, that is, authenticity in communication. The author emphasises that the natural use of two languages concomitantly shows the equal importance of both languages in tandem, which characterises the context as bilingual.

Picoli and Salomão (2020) investigate this principle in three teletandem dyads by observing the amount of time dedicated to both languages and moments in which there was code-switching, that is, during the time supposedly dedicated to one language, the other was used. Their results show that two dyads separated the languages equally while one used Portuguese for a longer time. The authors

conclude that to separate languages in terms of equal time does not mean mutual benefits of the practice.

We argue that both principles, reciprocity and separation of languages, underlie teletandem practice (together with the autonomy principle), but we acknowledge what Elstermann (2017, p. 31) states, i.e., “to associate the principle of reciprocity with the division of time in each language is a limitation”. As Picoli and Salomão (2020) conclude, separation of languages in telecollaborative contexts implies a monolingual perspective, which is not the case of teletandem as already pointed out by Lewis (2020). The authors also suggest that code-switching is not intrinsically related to the principle of separation of languages because this phenomenon is related to interlanguage and is motivated by other factors. They argue that “thinking that a bilingual should completely separate the use of two languages he/she knows seems to be a theoretical idealisation supported by a monolingual perspective, and not by the idea that linguistic knowledge of bilingual speakers is in contact and mutually influenced”² (Picoli & Salomão, 2020, p. 1621). They advocate that the improvement of the target language may depend partially on the time division of two languages, but it is also related to the expectations, objectives, and linguistic resources of the learners.

The literature in the field of Applied Linguistics reports several discussions related to language separation, especially in contexts of collaboration or telecollaboration. Amongst such works, we would like to highlight Dooly (2011), Schwienhorst and Borgia (2006), Kapec and Schwienhorst (2005) and Lee, Hill-Bonnet and Gillispie (2008).

Dooly (2011) reflects on perceptions and learning spaces. She analyses various learning processes based on extensive documentation, including teaching plans, messages posted on forums, participants’ responses, and a Wiki. She observed that the monolingual task was performed consistently, even when learners did not fully use the target language. This would occur because learners create a representation of the target language that is characterised by hybridity and bring traits of their native language.

The automatic system introduced by Schwienhorst and Borgia (2006) results from the authors’ perception that their students needed help keeping language separation during their Object-Oriented Multi-User Domain (MOO)³ interaction, where monitoring bilingualism needs to be more accurate. In such environments, the most proficient partner of an interaction tends to dominate communication. The same might also be relevant in other synchronous contexts, such as face-to-face and teleconferencing (our focus in this study).

The system developed by Schwienhorst and Borgia (2006) analyses each participant individually and synchronically to their interaction. The students were unaware of the arrangements, as the authors believed the awareness of a linguistic assessment system would negatively interfere with their language performance. On the one hand, such data compilation can bring more significant authenticity to the results; on the other, such a method can also raise ethical controversies, particularly concerning learners’ privacy.

Their unit of analysis was the speech turn. According to the authors, code-switching could negatively impact their results if the study had the lexis as its central unit. They compared the automatic segmentation results to a human classification of the segments. The objective was to measure the accuracy of the tool processing skills and the quality of its results. Schwienhorst and Borgia (2006) also automatically excluded all utterances shorter than ten characters long from the analysis and all ambiguous results regarding language definition.

Each session and its global statistics (including their bilingual proportion during each interaction) were analysed. Their results show that their system provides a relevant tool for students and instructors to control and understand bilingualism in tandem synchronous sessions. They also indicate that three out of six groups were imbalanced regarding language use.

Despite the fact that the study by Schwienhorst and Borgia (2006) is relevant for this research, as the authors report a successful experience of analysing the language separation principle in the context of tandem, some observations must be addressed. Firstly, the participants work with written interaction rather than oral. Even when using a MOO, there are nuances in face-to-face encounters through telematic tools that will undoubtedly introduce new variables to the analysis, especially interruptions. On the other hand, the authors excluded excessively short and potentially ambiguous interactions, simplifying the data classification process. Logically, interactions through synchronous text have clear markers, such as line breaks. However, a conversational turn can be broken into small lines of interaction, posing problems in segmentation and excluding short utterances, which would be part of a longer turn.

Kapac and Schwienhorst (2005) studied learners' response to the use of the same tool, as well as their perception of teaching, learning and bilingualism aspects in teletandem practices. The authors used a digital questionnaire based on numerical rating scales. Their results show a relative contradiction in the perception of the target language use. The learners' comments evaluated the tool either as positive or showed indifference. Lee, Hill-Bonnet and Gillispie (2008) qualitatively study the principle of language separation in a Spanish-English bilingual teaching context. Their database includes videos of recorded interactions, field notes and spontaneous interactions. They observed that code-switching occurred naturally during interactions, especially outside the classroom, without negatively influencing the learners' performance.

Despite their different approaches, these studies are relevant for developing methods for analysing language separation and reciprocity. Their results, in general, show that using some words or expressions not in the target language does not seem to jeopardise the learning process.

In this paper, we analyse 16 Initial TOS from different pedagogical scenarios to verify whether and how participants rely on the principle of language separation. We assume that, in the first encounters, participants will try to respect the principles as they are told to do so when they apply to take part in the project.

Enrolling in a teletandem project implies that the individual will commit to participate in a specific number of Teletandem Oral Sessions (TOS) and mediation sessions i.e., a meeting that occurs after TOS between the teacher/mediator and the participants during which aspects of the interactions are discussed, goals are evaluated, and new paths are discussed. TOSs are one of the macro-tasks (together with the mediation session) responsible for the synchronous language and culture exchange between participants. Other tasks apply to different learning scenarios, i.e., the expected pedagogical organisation of cohorts as well, but the common core of each and every teletandem project is the two macro-tasks (Aranha & Leone, 2017).

Aranha and Leone (2017) state that the TOS is part of a learning scenario, together with the mediation sessions. During each TOS (on average, a cohort participates in eight TOSs over a semester), students develop linguistic, cultural, pedagogical, and digital competencies while they learn the language of the other and help the partner. The TOS characteristics within teletandem learning scenario (TTLS) at UNESP (Universidade Estadual Paulista) include: (i) the conversation in two languages following the principles of reciprocity, autonomy and separation of languages; (ii) the exchange of texts written in both languages, corrected and discussed by the most proficient partner; and (ii) the writing of a reflective diary after each TOS.

The first 15 minutes of Initial Teletandem Oral Sessions (iTOS) have been studied by Aranha (2014) and Rampazzo and Aranha (2018, 2019). The hypothesis presented by Aranha (2014) is that initial sessions are part of a system of genres that have shared characteristics and that such a genre could be recognized in terms of rhetorical actions. In her work, she analyses 10 initial sessions from different learning scenarios and finds recurrent moves in rhetorical, content, and linguistic terms. Rampazzo and Aranha (2018, 2019) expand the analysis and reach further results. The authors reinforce the thesis that in the first 15 minutes, the discourse in iTOS is organised in rhetorical parts that are similar in content in each sample analysed. However, they find out that “ (...) although very similar, the organisation of the first 15 minutes of the iTOS may vary depending on the learning scenarios (...)” (Rampazzo & Aranha, 2019, p. 17). The authors also emphasise that “despite the different learners involved in the partnerships and interactions, students in teletandem organise their discourse in the beginning of the initial oral session in similar manners” (Rampazzo & Aranha, 2019, p. 24). Their theoretical background is composed of studies on genres, specially from New Rhetoric and Socio-Rhetoric backgrounds, so they also affirm that the iTOS seems to be going through a process of standardisation as participants understand this situation as similar to others and start using certain rhetorical responses, obtained through the use of specific textual characteristics. Following the study of the first 15 minutes, Rampazzo (2021), who analyses the full-length of iTOS, corroborates the thesis that it is a telecollaborative genre developed and appropriated by learners as they engage in the tasks recommended by their professors. The author argues that the organisation of iTOS in its full-length is

similar, although also fluid and malleable depending on the characteristics of the pedagogical scenarios.

We argue, following Rampazzo and Aranha (2018) and Telles (2015), that teletandem oral sessions are not a simple chat and, although the session may have traces of informal conversations between friends, participants also negotiate meanings, refer to the tandem principles they are supposed to follow, clarify linguistic usage and meanings, improve cultural understanding, share information and knowledge, learn how to collaborate and, ultimately, have some fun.

This paper aims at answering the following research questions: 1) Do the participants during iTOs respect the language separation principle, and if so, how? and 2) To what extent do lexical choices reflect this principle?

2. Methods

The corpus results from 16 Teletandem Initial Sessions (iTOS), all including Portuguese and English partnerships. Data are from MulTeC (Multimodal Teletandem Corpus) (Aranha and Lopes, 2019), which comprises information from 282 university Brazilian and American students (Portuguese/English) of 16 cohorts enrolled in different majors. The corpus compiles TOS, reflective diaries, initial and final questionnaires, chat entries and diverse genres of texts written in English and exchanged with the partners. All Portuguese speakers (PS) are Brazilian citizens (the majority Language and Literature or Translation Studies at UNESP), while English speakers (ES) are US citizens, some from Spanish background. Every partnership in an iTOS is supposed to be performed in two languages for the same amount of time. Participants choose either Portuguese or English to start the session with and, after half an hour, they change to the other. Each of these moments is referred to in this study as Portuguese or English subsection.

This iTOS configuration has motivated the way our corpus is organised. All iTOSs were broken into Portuguese and English subsections according to the speaker's language proficiency. As a result, this study compares four subcorpora:

Table 1: Corpora and Subcorpora

Portuguese sessions (main corpus 1)	English sessions (main corpus 2)
Lexical choices by Portuguese speakers (PS) during Portuguese sessions (subcorpus 1)	Lexical choices by Portuguese speakers (PS) during English sessions (subcorpus 3)
Lexical choices by English speakers (ES) during Portuguese sessions (subcorpus 2)	Lexical choices by English speakers (ES) during English sessions (subcorpus 4)

The next step was to make wordlists for each subcorpus, using text processing R (R Core Team, 2021) packages such as Tidytext (Silge & Robinson, 2016), responsible for processing and counting words, and Dplyr (Wickham et al., 2021), responsible for data manipulation. These wordlists were saved in CSV (Comma Separated Values), a standard text-only table format, and later edited

in a conventional spreadsheet application. Each word was classified in a second column according to the language:

1. ING: represents words in English
2. OTH: short for others, represent languages with very small number of occurrences, such as Italian and French
3. POR: represents words in Portuguese
4. SPA: represents words in Spanish
5. COG: cognates, words similar in Portuguese and English
6. UND: short for undefined, includes interjections, words that characterise paralinguistic expressions and the coded name of participants.

Cognate words (COG) mean a particular category due to their possible contribution to mutual comprehension during the sessions. It is widely known that ordinary English courses and ESP reading courses in Brazil (Celani, 2005; Ramos, 2008) have emphasised the importance of cognate recognition as a strategy for comprehension, making it a common ground strategy for many learners.

Undefined (UND) is a different category due to the methods used in transcription. The transcripts of the Teletandem Oral Sessions were carried out using the Transana software, developed for the transcription and analysis of textual data. The software is installed on a computer, and the person who transcribes the session is responsible for uploading the media - video (mp4) or audio (mp3) - and using the tools provided by the software while transcribing. The transcripts favour verbatim transcription of the archives and follow MulTec protocols. However, during the analysis, we noticed the transcription of interjections did not follow any specific system, most of their translation into written language depended mainly on each transcriber's interpretation; each transcriber seemed to use their own pattern - which was not necessarily consistent throughout the transcription. Since we did not intend to interfere in the original transcription to normalise possible inconsistencies, the solution of creating a specific category seemed more suitable. It is also important to point out that such an issue was present only in this type of lexis, which seemed not to interfere consistently with the results. This category also embraces transcription codes for usernames (for example, 2012_I8F4_UGA1i), which follow a protocol established for the corpus compilation and does not support language specific features. Other Languages category (OTH) includes languages with very few lexical occurrences, such as Italian and French. Due to their low numbers, such languages were not considered in the quantitative or qualitative analyses.

We recycled classifications amongst subcorpora, using standard R list comparison (R Core Team, 2021). In other words, lexical items in *subcorpora 1* were classified subjectively according to the criteria mentioned above, then they were automatically applied in the classification of the lexical items of *subcorpus 2* and vice-versa, cumulatively throughout the subcorpora. The principle here is to save time for the classification of already consolidated words. A pronoun like

Eu (I), for example, will always represent a Portuguese lexical choice, whichever subcorpora it is. The remaining and exclusive lexis of each subcorpus was manually classified. The next step was to observe whether the difference in lexis was statistically meaningful in each subcorpus.

Lexical choices in the Portuguese and English segments in each iTOS were compared using Analysis of Variance (ANOVA). The research hypothesis predicts that such groups are different in terms of the lexis they use in each language. Each word had its frequency taken into account individually in order to properly weigh the language it represents in each subcorpus.

3. Results and discussion

Table 2 shows the lexical choices made by PS during Portuguese subsections. Its layout is to be followed throughout all data analysis; the first column brings information regarding the languages used by such speakers, always in alphabetical order, followed by the number of types and tokens for each language and by the type/token ratio — a standard measurement of lexical diversity. The last two columns represent the normalised types and tokens information, offering a more precise means of comparison.

Table 2: *Portuguese Speakers — Portuguese subsections (subcorpus 1)*

Languages ^a	Types	Tokens	TTRatio ^b	NTypes ^c	NTokens ^c
COG	48	650	0.07	21.3	41.29
ING	566	3203	0.18	251.11	203.44
POR	1606	11222	0.14	712.51	712.78
SPA	7	9	0.78	3.1	0.57
UND	27	660	0.04	11.98	41.92
Total	2254	15744	1.21	1000	1000

^a COG: Cognates, POR: Portuguese, ING: English, SPA: Spanish, UND: Undefined

^b TTRatio: Types/Tokens

^c Normalised by 1000

Portuguese is the most common lexical choice (713 per thousand), followed by English (251 per thousand); cognates and undefined have lower numbers (21 per thousand and 12 per thousand, respectively). Though Spanish is the least used language, it has a high types and tokens ratio. However, if one takes a closer look at the normalised columns, one would notice that Spanish has very low occurrences per thousand words. This was common to all languages that were less frequently present in the subcorpora (see discussion on other subcorpora), which probably means that the types/tokens ratio is not valid for low-frequency lexis in a corpus.

We may suggest that the use of Spanish is motivated by the similarity between both languages. As Portuguese speakers, we constantly observe the similarity of

both languages as a common strategy for coping with Portuguese vocabulary. Besides, some of the ES partners have had some previous Spanish contact or are from Spanish background (Ex. 1).⁴

Ex. 1

- ES sí... ah, eu gosto muito de a lengua a portuguesa.
yes, I like the Portuguese Language a lot
- PS Ah sim... yo hablo un poquito en espanhol también
Oh yes, I speak a little of Spanish as well

Portuguese and English have a similar type/token ratio. Although the former is more commonly present than the latter, PS tend to repeat lexis equitably in these languages. If we compare the three more frequent languages, Portuguese is three times more common than English as a choice, while English is eight times more frequent than the cognates, which have a 10% repetition ratio. This result is within our expectations since the use of cognates by PS seems to be the result of lexical similarity in both languages. Words with Latin roots have been commonly present as a language comprehension and communication strategy in EAP, ESP and ESL courses in the Brazilian context (Celani, 2005). Teachers highlight that Portuguese Latin origins can be used in their favour, encouraging learners to use such words in their speech and to associate those which are of Latin origin in English with their meanings in Portuguese. In ex. 2, the PS is trying to explain to his/her ES partner a common Brazilian dessert (*brigadeiro*). As the ingredients are listed, the lexis *chocolate em pó* (chocolate powder) is not translated into English, as the PS partner relies on its similarity to switch back into Portuguese. The same strategy seems to be present in example 3.

Ex. 2 (...) esse doce vai condensed milk condensed milk ... esse doce .. e vai chocolate também em pó sabe?

In ex. 3, the PS partner is explaining the meaning of *a gente*, an informal register of *nós* (we) in Brazilian Portuguese.

Ex. 3 (...) that *a gente*⁵ would be **informal** not **formal** (...)

Most Teletandem studies about TOSs focus on other aspects of communication, such as the rhetorical structure of such texts as genres (Rampazzo & Aranha 2018; 2019). Only recently has the Brazilian Teletandem initiative been compiled into a corpus (Aranha & Lopes, 2019), with the possibility of qualitative/quantitative research comparing more significant amounts of data. Using lexical choices across the sessions is quite a new approach to studying language separation. As a result, it brings a more consistent proposal to question an important myth regarding using languages other than the target ones. As we will observe in the analysis, using languages other than the target does not consist of a “deviation” from the session objectives, but a valid communication strategy

that emerges when the common ground in the target language is not established. Such use reiterates the importance of keeping the interaction going either in English or Portuguese. Participants resume the use of the language as soon as the communication difficulties are overcome.

The use of English vocabulary takes place in two situations. The first is a code switch that favours the comprehension of an ES (ex. 4). As we see in ex. 4, the PS partner switches from Portuguese into English in the first turn, and in the fifth, as the ES partner shows some difficulties in understanding the question about the number of possible languages to choose. The second results from the first. As such code-switching takes place, teletandem partners tend to keep on speaking English and, sometimes, it takes several turns until Portuguese is resumed (ex. 4).

Ex.4

- PS e você é you can choose the language?
 ES no
 PS how many languages do you have? in University? what languages can you choose?
 ES ahm?
 ES ah... so many
 PS like Portuguese French Spanish?
 ES francês alemangi ale/alemanha? alema
 PS alemão
 ES alemão... zulu
 PS that's too many languages
 ES sim

Interjections work as a confirmation of meaning or as validation (ex.5). As the participants signal they understand each other, this helps the speaker to keep talking, a confirmation they are on the right communication path.

Ex.5

- ES é mas también se você quer é ver é a gramática... você pode por exemplo é ler jornais acadêmicos
yes but also if you want to study grammar... you can for example read newspapers
 PS uhum

The results for ANOVA (table 3), comparing language choices for PS during Portuguese subsections is significant in their variation ($F(4, 2249) = 3.97, p = .003$) in a 95% confidence level. The DIFF column estimates the group mean difference, the LWR and UPR columns provide the lower and upper confidence interval bounds on such a difference. One might suggest that Portuguese is more frequent because it might fulfil the purpose of speaking that language during the session, while the presence of languages other than Portuguese might be

motivated by the need for mutual comprehension. Possibly such use takes place when an ES does not seem to have the necessary command of Portuguese. As a result, either ES switches to a language she/he feels more comfortable with, or the BP switches to English.

Table 3: ANOVA results for Subcorpus 1

	DIFF ^a	LWR	UPR	p [*]
ING-COG	-7.88	-18.94	3.17	0.29
POR-COG	-6.55	-17.32	4.22	0.46
SPA-COG	-12.26	-42.01	17.49	0.79
UND-COG	10.90	-6.79	28.59	0.44
POR-ING	1.33	-2.27	4.92	0.85
SPA-ING	-4.37	-32.34	23.59	0.99
UND-ING	18.79	4.30	33.27	0.00
SPA-POR	-5.70	-33.55	22.15	0.98
UND-POR	17.46	3.19	31.73	0.01
UND-SPA	23.16	-8.03	54.35	0.25

^aDIFF=difference, LWR, =l ower, UPR= upper

^bCOG: Cognates, POR: Portuguese, ING: English, SPA: Spanish, UND: Undefined

^{*}F(4, 2249) = 3.97, p = .003

Table 4 shows the lexical choices made by ES during Portuguese subsections. Portuguese is the most frequent language (592 per thousand tokens), followed by English (251 per thousand), undefined (104 per thousand), cognates (52 per thousand) and Spanish (23 per thousand). If we compare tables 2 and 3, ES proportionally uses less Portuguese and more English than PS. Besides, PS and ES tend to have a slightly more diverse use of English (0.20 TTRario) than Portuguese (0.17 TTRario).

Table 4: English Speakers — Portuguese subsections (subcorpus 2)

Language ^a	Types	Tokens	TTRatio ^b	NTypes ^c	NTokens ^c
COG	56	332	0.17	25.93	25.81
ING	649	3235	0.2	300.46	251.52
POR	1286	7620	0.17	595.37	592.44
SPA	109	296	0.37	50.46	23.01
UND	40	1346	0.03	18.52	104.65
OTH	20	33	0.61	9.26	2.57
Total	2160	12862	1.55	1000	1000

^aCOG: Cognates, POR: Portuguese, ING: English, OTH: Other Languages, SPA: Spanish, UND: Undefined

^bTTRatio: Types/Tokens

^cNormalised by 1000

Table 5 shows the results for ANOVA, comparing language choices for ES during Portuguese subsections. The results of ANOVA are significant ($F(5, 2154) = 8.98, p < .001$) in a 95% confidence level. The DIFF column estimates the group mean difference, the LWR and UPR columns provide the lower and upper confidence interval bounds on such a difference. These results are consistent with the PS in this subsection, as both groups of speakers tend to use Portuguese more significantly and consistently, switching to other languages when some communication problem arises. If we compare these numbers to the earlier analysis, such figures might help us understand the use of those languages. English is proportionally a more common choice, while there is a slightly higher use of cognates. So, to cope with their difficulties in using Portuguese, it appears that ES tends to intensify the use of a better-known language, look for cognates and seek similarities with other Latin languages. These strategies, however, seem to occur without compromising the role of Portuguese as the primary language in these subsections. Such results seem to reflect a strategy for language use and learning.

Table 5: ANOVA results for Subcorpus 2

	DIFF ^a	LWR	UPR	p [*]
ING-COG	-0.94	-11.68	9.80	1.00
OTH-COG	-4.28	-24.36	15.81	0.99
POR-COG	0.00	-10.53	10.52	1.00
SPA-COG	-3.21	-15.89	9.46	0.98
UND-COG	27.72	11.76	43.68	0.00
OTH-ING	-3.33	-20.84	14.17	0.99
POR-ING	0.94	-2.77	4.65	0.98
SPA-ING	-2.27	-10.25	5.71	0.97
UND-ING	28.67	16.10	41.23	0.00
POR-OTH	4.28	-13.10	21.65	0.98
SPA-OTH	1.07	-17.69	19.82	1.00
UND-OTH	32.00	10.88	53.12	0.00
SPA-POR	-3.21	-10.90	4.48	0.84
UND-POR	27.72	15.34	40.10	0.00
UND-SPA	30.93	16.68	45.19	0.00

* $F(5, 2154) = 8.98, p < .001$

^aDIFF=difference, LWR, =lower, UPR= upper

As we have discussed earlier, the motivation for such results is that the use of English by ES is related either to code-switching as a strategy for unknown vocabulary or miscomprehension (ex.4 and ex.6) or to the maintenance of English after code-switching (ex.4).

Ex. 6

(...) how do **you** say so in Portuguese like (...)

(...) mas eu ouvi **like or** como.

The presence of English during the Portuguese subsections seems to be motivated by the learning needs of ES, so the use of English vocabulary results in their need to establish a common ground for the interaction or to avoid communication breakdown and misunderstandings. This sort of language use might be defined as a cooperation strategy, as one speaker notices the need for the code-switching, she/he complies by helping to establish the necessary mutual understanding. The initial code-switching makes a change into the section, which temporarily switches into a more comfortable means of interaction to the learner. As the interaction difficulties ease, Portuguese use recommences.

The use of Spanish seems to be related to phonological similarities, such as *mucho*, taken for *muito* (Ex. 7) or *entonces*, instead of *então* (Ex. 8). In some cases, like ex.7, it is possible to observe that such confusion is commonly made by ES, who tend to think of Portuguese and Spanish as similar languages.

Ex. 7 (...) quando eu intentava a aprender Inglês foi **mucho** difícil
 (...) *when I tried to learn English it was very difficult*

Ex. 8 (...) difícil quando o quando eu eu tome português para um ano é e **entonces** não tem para uma (...)
 (...) *difficult when I took Portuguese lessons for a year so there is not (...)*

Ex. 9 (...) é **mio mio** medio nome ma (...)
 (...) *it is my my name (...)*

The use of cognates is related to familiar words in both languages. In example (10) a ES uses the word *cinema* as a common lexis in both languages, while in example (11) the cognate is the name of a Central America country.

Ex. 10 (...) oh ! eu queria falar ou conversar com você sobre seu ah . . . love for **cinema** and translation sim porque é interessante o ?
 (...) *oh! I would like to talk to you about ah... love for cinema and translation yes because it is interesting to (...)*

Ex. 11 (...) hay hay mais gente em São Paulo que todo ah **Honduras** (...)
 (...) *hay hay more people in São Paulo than talvez Honduras (...)*

Table 6 brings the language distribution of subcorpus 3: Brazilians during English sub-sections. English is the most used language (846 per thousand), followed by Portuguese (63 per thousand) and undefined (50 per thousand). Cognates (40 per thousand) and Spanish are the least frequent. These results show that Brazilians tend to use English more frequently during these subsections, something expected during a tandem practice and supported by the principle of language separation. Nevertheless, if we compare the Type/Token ratio, the use of English tends to be less diverse than the use of Portuguese. In other words, PS uses more English than Portuguese during their English subsections, but they

tend to repeat the same words throughout the sessions. So, Portuguese (0.33 TTRatio) is less frequent, but tends to be richer in terms of lexical choices.

Table 6: *Portuguese Speakers — English subsections (subcorpus 3)*

Language ^a	Types	Tokens	TTRatio ^b	NorTypes ^c	NorTokens ^c
COG	75	682	0.11	43.33	40.6
ING	1253	14213	0.09	723.86	846.06
OTH	3	3	1	1.73	0.18
POR	352	1061	0.33	203.35	63.16
SPA	8	8	1	4.62	0.48
UND	40	832	0.05	23.11	49.52
Total	1731	16799	2.58	1000	1000

^aCOG = Cognates, POR = Portuguese, ING = English, OTH = Other Languages, SPA = Spanish, UND = Undefined;

^bTTRatio= Types/Tokens;

^cNormalised by 1000

The results for the ANOVA test for subcorpus 3 (table 7) it are significant in their ($F(5, 1725) = 3.00, p = .011$) in a 95% confidence level. The DIFF column estimates the group mean difference, the LWR and UPR columns provide the lower and upper confidence interval bounds on such a difference.

Table 7: *ANOVA results for Subcorpus 3*

	DIFF ^a	LWR	UPR	p [*]
ING-COG	2.25	-11.58	16.08	1.00
OTH-COG	-8.09	-76.59	60.40	1.00
POR-COG	-6.08	-20.87	8.72	0.85
SPA-COG	-8.09	-51.36	35.18	0.99
UND-COG	11.71	-11.07	34.48	0.69
OTH-ING	-10.34	-77.59	56.91	1.00
POR-ING	-8.33	-15.35	-1.31	0.01
SPA-ING	-10.34	-51.61	30.92	0.98
UND-ING	9.46	-9.23	28.14	0.70
POR-OTH	2.01	-65.44	69.47	1.00
SPA-OTH	0.00	-78.76	78.76	1.00
UND-OTH	19.80	-49.84	89.44	0.97
SPA-POR	-2.01	-43.61	39.58	1.00
UND-POR	17.79	-1.63	37.20	0.09
UND-SPA	19.80	-25.26	64.86	0.81

* $F(5, 1725) = 3.00, p = .011$

^aDIFF=difference, LWR, =lower, UPR= upper

English is spoken most frequently in this subcorpus, and its means of occurrence show such frequency does not tend to be by chance. English is possibly the most frequent language as it might fulfil the purpose of the learning situation. The presence of languages other than English takes place in three situations: 1) vocabulary questions (ex. 12); 2) hesitations (ex. 13) and proper names (ex. 14). Vocabulary questions (ex. 12) are a way to learn new words in English; a PS usually interrupts and says a word in Portuguese, trying to get its equivalent in English. This sort of strategy relies on the ES' Portuguese knowledge, which is facilitating the learning process. During hesitations (ex. 13), PS occasionally switches back to Portuguese. However, the PS does not actually seek to establish a dialogue in Portuguese; it happens when PS is thinking or organising their discourse. Finally, words like *letras* (ex. 14) reflect the official name of the major at UNESP. Such names usually are not to be translated during the interactions.

Ex. 12

Yes when I was a child I didn't ' like so much be an only child but now I I like I how I can say eu estou **acostumada** ?

Ex. 13 yes ... ahn...yeah ... yes but ... é the ... the English is the most most important ?

Ex. 14

If you want study ahn **Letras** you can

The use of cognates seems to be related to a strategy for coping with difficulties in vocabulary during TOSs. They are associated with common lexis of both languages, mostly proper names, like *Carnaval* (ex. 15), and to words with similar meaning in both languages, like *cinema* (ex. 16). Brazilians usually use cognates as a strategy for communicating vocabulary difficulties.

Ex. 15

I'm Brazilian and I don't like carnaval (...)

Ex. 16

you like to go to the cinema (...)

Other languages besides English and Portuguese are sporadic and not statistically meaningful. Mostly, they happen in hesitation moments, apparently to facilitate the interaction with the ES partner when seeking for proper vocabulary.

Ex. 17

I don't know but do you understand? It's very difficult... for me... I love
PS live with my parents but sometimes my mother is very protec I don't
know... protetora... he protect

ES Protective

PS Yeah. He protects a **lo a los** me so this is difficult for me

The languages spoken by ES speakers during English subsections are presented on Table 8. English is the most frequent language (870 per thousand tokens) followed by Portuguese (54 per thousand) and cognates (24 per thousand); Spanish has the lowest frequency.

Table 8: *English Speakers — English subsections (subcorpus 4)*

Language ^a	Types	Tokens	TTRatio ^b	NorTypes ^c	NorTokens ^c
COG	54	436	0.12	26.76	23.94
ING	1616	15860	0.1	800.79	870.85
OTH	2	3	0.67	0.99	0.16
POR	307	981	0.31	152.13	53.87
SPA	9	18	0.5	4.46	0.99
UND	30	914	0.03	14.87	50.19
Total	2018	18212	1.73	1000	1000

^aCOG = Cognates, POR = Portuguese, ING = English, OTH = Other Languages, SPA = Spanish, UND = Undefined;

^bTTRatio= Types/Tokens;

^cNormalised by 1000

Table 9 displays the distribution of types and tokens across subcorpus 4. The results for ANOVA are significant in a 95% confidence level ($F(5, 2012) = 3.49, p = .004$).-. As we compare subcorpus 3 and 4, it may be observed that English is the most common lexical choice. This result is likely to be consistent with their primary task: use English for communication during the second half of the iTOS. Portuguese and English normalised number of tokens is similar across subcorpus 3 and 4. ES partners tend to use English a bit more than PS, who, on the other hand, favour Portuguese a bit more. However, such a difference is too faint to be statistically meaningful, so it is probably a result of PS 'tendency to move into Portuguese, as some vocabulary doubt arises. These results indicate that such linguistic choices were not made by chance. The difference appears to be motivated by their commitment to the language separation principle.

Table 9: *ANOVA results for Subcorpus 4*

	DIFF ^a	LWR	UPR	p [*]
ING-COG	1.74	-13.45	16.94	1.00
OTH-COG	-6.57	-85.67	72.52	1.00
POR-COG	-4.88	-21.09	11.33	0.96
SPA-COG	-6.07	-45.62	33.47	1.00
UND-COG	22.39	-2.62	47.40	0.11

OTH-ING	-8.31	-86.03	69.40	1.00
POR-ING	-6.62	-13.46	0.22	0.06
SPA-ING	-7.81	-44.53	28.90	0.99
UND-ING	20.65	0.41	40.89	0.04
POR-OTH	1.70	-76.23	79.62	1.00
SPA-OTH	0.50	-85.37	86.37	1.00
UND-OTH	28.97	-51.25	109.18	0.91
SPA-POR	-1.20	-38.34	35.95	1.00
UND-POR	27.27	6.26	48.28	0.00
UND-SPA	28.47	-13.28	70.21	0.37

* $F(5, 2012) = 3.49, p = .004$

^a DIFF=difference, LWR, =lower, UPR= upper

Portuguese is preferred during two situations. In the first, the ES partner asks the Brazilian to express him/herself in Portuguese to clarify a specific vocabulary question. In these situations, the ES keeps speaking Portuguese, sometimes switching to English and back to Portuguese, always checking if the meaning they are co-constructing is the most appropriate. Example (18) brings a piece of interaction that represents such a situation. The ES partner (ES) asks the PS partner (PS) to express his/her question in Portuguese. PS misunderstands the question, and finally expresses himself in Portuguese (ES). As (ES) finally understands what (PS) means, they resume speaking English.

The topic is related to Brazil and the ES partner seizes the moment to learn some Portuguese vocabulary, while using English for the most part of the interaction. Example (19) shows a discussion about the equivalence of the Portuguese word for the English *mountain*. The interaction keeps flowing in English, as the ES partner seems to have some difficulties due to the phonetic similarity of *morro* (lower mountains) and *poco* (little in Spanish). The PS emphatically corrects such phonetic misunderstanding

Ex 18

- ES ok what are you trying to say in that sentence please say again can you tell me in Portuguese for a second what you are trying to say to tell me to explain you what the translation in English...
- PS Ah I don't...
- ES Ah você você é você pode é dizer me dizer é o que você é queria é dar me dar a entender... é em português?
Ah you you are you can say that you meant you meant, is it Portuguese?
- PS Isso sobre cidade?...
- PS *About the city?*
- ES Just la sentence about the tourist

- Ah sim é em... o Rio de Janeiro é... é um importante centro comer centro de turismo mesmo... é ele é conhecido pelo pelo turismo... e e São Paulo que eu tinha te falado era que São Paulo é mais importante nessa questão comercial de indústrias
- PS *Oh yes, it's in... Rio de Janeiro is... it's an important centre to eat well, a tourism centre... it's known for tourism... and São Paulo that I told you was that São Paulo is most important in its commercial question of industries*
- ES *uhum] e é a capital financeira vamos se dizer assim*
- ES *uhum] it is the financial capital, let's say*
- ES *ok... hum... interesting interesting... yeah the thing is when you want to say for example ce centro turístico is tourist centre*

Ex 19

- ES *ah... eh... Pão de Açúcar that you cannot climb but it looks good for the city but they're not ... extremely high*
- PS *yes that's the thing we don't have ah... extremely high moun mountains we have some some what's the name? morro? in English*
- ES *what's that?*
- PS *morro do you know? morros it's kind MORROS it's like ah?*
- ES *pocos*
- ES *pocos?*
- PS *MORROS with with R morros*
- ES *morros*
- PS *ah it's like mountains but not that high*
- ES *oh like small hills like outeiro like that?*

The use of words common to both languages commonly bring names of cities/countries/regions (ex 20, 21), or words which have a similar origin (ex 21).

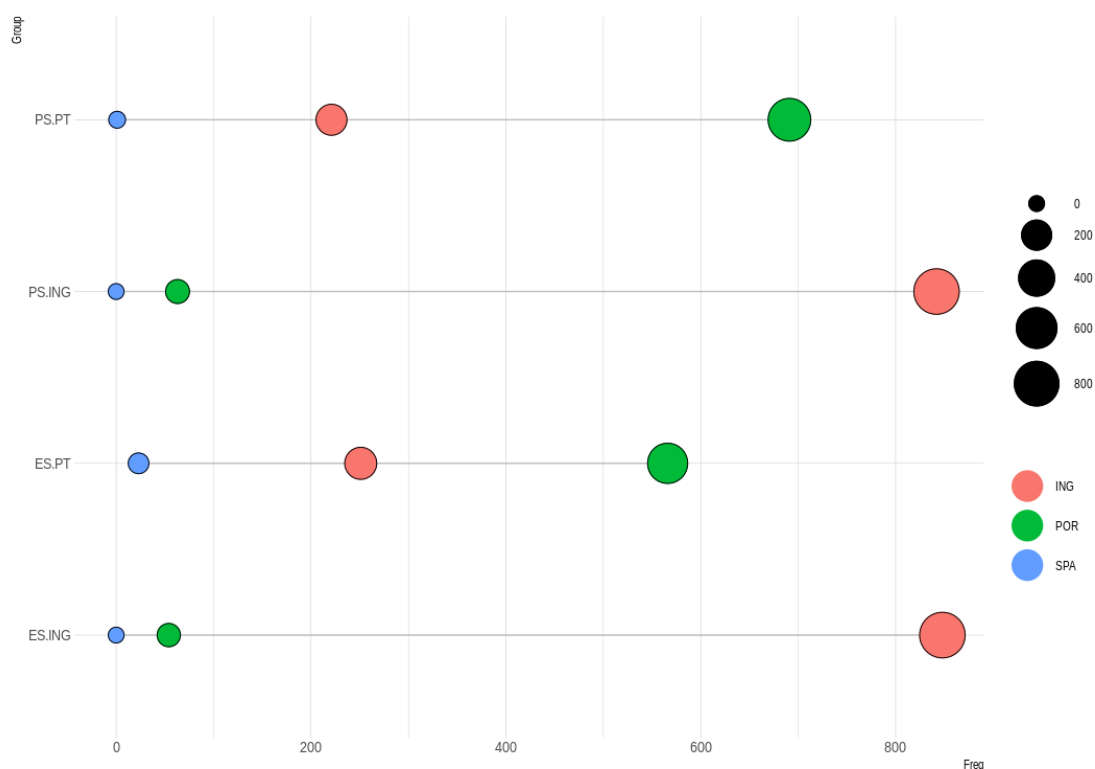
Ex 20

ES	<i>ah yeah that's true in Peru we don't have to many difference neither... we say city as the capital of every province...</i>
PS	<i>yes and city Lima for example even Lima is huge compared to my little city that is like that yeah...</i>

Ex 21

what it's called but it's very popular in in the o sertão ?

Figure 1 displays a comparison amongst all languages across the four subcorpora. The colours represent each language, the lines bring information about group and subsection, and the size represents the number of occurrences normalised by 1000. One would notice that during Portuguese and English subsections, both Brazilian and Non-Brazilian tend to use the language assigned, expected for that part of the SOTi .

Figure 1: Comparison of frequencies across subcorpora

Note: PS.PT = PS subsections (subcorpus 1), ES.PT = ES – Portuguese subsections (subcorpus 2), PS.ING = PS – English subsections (subcorpus 3), ES.PT = ES – English subsections (subcorpus 4)

English plays a more critical role in the Portuguese subsections than Portuguese in the English subsections. There is a slightly higher use by non-Brazilians, possibly a result of the strategies we have identified during the analysis of subcorpus 2, meaning that non-Brazilians tend to use English to navigate during their language issues when speaking a foreign language more often than Brazilians try to speak Portuguese during their English parts.

In this section, we surveyed the lexical choices made by the participants in the iTOS. Our results show that the use of the target language occurs consistently in higher and significant proportion; however, other languages emerge when there are communication problems to be resolved. In such cases, the most common tendency is to seek support in the native language of the participant, in this case, Portuguese and English. The target language tends to return when such difficulties are overcome, which can be interpreted as the participants' commitment to maintaining the principle of language separation.

4. Final Remarks

This article studied 16 initial teletandem sessions to observe whether learners followed the separation of language principle and, if so, how. First, we described the context by emphasising the principles that underlie teletandem practice. We also discussed some research that has already investigated them from qualitative and quantitative perspectives.

The iTOS were segmented into their Portuguese and English subsections. We analysed the lexical choices, identifying the language performed by Portuguese and English native speakers. We carried out linguistic classification based on wordlists, which were recycled and used cumulatively. Scripts in the R programming language performed wordlists, statistical tests, and surveys of examples. The ANOVA statistical test was applied to verify the lexical variance during each section - Portuguese and English.

Regarding the first research question, the results show that the principle of language separation tends to be respected. The cross-use of languages (English during the Portuguese subsection and vice versa) occurs when there are questions about vocabulary, phonetics misunderstanding of words or requests for clarification by one of the participants. ES speakers tend to use more English during Portuguese subsections, while PS use less Portuguese in English subsections. Most of the time, the use of English and Portuguese reinforces Lewis' (2020) statement about tandem being a bilingual language environment. Other languages also occur during the interaction, and the use of Spanish seems a bit higher due to the origin of the US participants. However, such use is sporadic and not statistically relevant. Their presence seems to be motivated by phonetic difficulties of ES.

Regarding the second research question, this article contributes to the study of teletandem, primarily by analysing the linguistic data of initial sessions, making it possible to understand some of the participants' language use. Although there is a cross-use of languages in the subsections, or even a greater diversity of languages than we expected, the motivation for such a phenomenon is strategies for meaning, mutual understanding and learning new vocabulary. Thus, we could not find evidence to corroborate the idea that PS or ES suspend the target language during the subsections.

When we observe the result of this research in comparison with others, we observe that it contributes to understanding how the principles of separation actualise in synchronous teletandem video sessions.

Like Lee, Hill-Bonnet and Gillispie (2008), our study showed that the use of non-target languages was a kind of tapestry, responsible for promoting the continuity of interaction when the use of English or Portuguese would make it difficult. The results also show that this change could indicate the learners' commitment to such principle since they favour learning and knowledge sedimentation by accepting the proposed language change. The resumption of the target language in the interaction seems to reaffirm the commitment to the

principle of separation, demonstrating, like Dooly (2011), that the use of other languages does not belittle the performance of a monolingual task.

As discussed earlier, Schwienhorst and Borgia (2006) evaluate the performance of an automatic linguistic classification system applied to tandem. Differently from this study, the researchers use the speech shift as the unit of analysis. Their system needs a segment of, at least, five words, which naturally excludes small turns. Several packages within the R programming language perform this type of automatic analysis, also requiring larger textual units. However, this approach would make other languages, such as Spanish, even less frequent than the results showed. Such language identifiers capture the words in a segment and identify its language, then classify the segment by most of the regular lexicon. Our semi-automatic approach entailed a delicate control of the lexis used by learners, allowing us to spot specific uses and perform a more refined interpretation of the role of these other languages.

Statistical analysis established the significance of the use of languages in each of the sections. Although languages such as Spanish and Italian are present, their frequency does not justify the perception that they would jeopardise the process of interaction in the target language. Instead, they support the building of knowledge represented by English or Portuguese. Second, the analysis showed that the difference between the averages of use is not at random; the learners are effectively using Portuguese and English as main languages at the appropriate moments of the teletandem session. Learners seem to build a trans-linguistic space (Canagarajah, 2020) that serves the target language learning process.

Finally, this study has some limitations. This article could be expanded regarding the number of initial sessions analysed. The methodology we have developed could be applied to the other sessions to elicit their strategies and verify possible regularities across different moments of teletandem practice. These data should also be analysed from a trans-linguistic perspective, including multimodal elements.

Acknowledgements

Rodrigo Esteves de Lima Lopes would like to thank CNPq (#311099/2021-1) and Solange Aranha would like to thank FAPESP (#2016/18705-9 and # 2019/14271-2) for the financial support for this research. Both authors would like to thank Drs. Suzi Spatti Cavalari, Claudia Zavaglia and Laura Rampazzo for her critical reading and valuable remarks on earlier versions of this paper.

Notes

1. In this paper, both terms will be used to refer to the same foreign language learning practice.
2. From the original: “pensar que o bilíngue deve separar completamente o uso das duas línguas que conhece parece ser uma idealização teórica que se respalda em um paradigma monolíngue, e não na ideia de que os conhecimentos linguísticos dos falantes bilíngues estão em contato e se influenciam mutuamente (Picoli & Salomão, 2020, p. 1621).

3. Object-Oriented Multi-User Domain (MOO) are online text-based virtual environments where multiple users are connected and interacting simultaneously.
4. All examples will be translated from Portuguese into English. However, some phonetic details, mainly the use of Spanish words in Portuguese communication are impossible to translate.
5. A gente is an informal variation of nós (we), commonly present in oral interactions.

References

- Aranha, S. (2014). Os gêneros na modalidade de teletandem institucional integrado: a primeira sessão de interação. In: E. L. NASCIMENTO, & R. H ROJO.. (Org.) *Gêneros de texto/discurso e os desafios da contemporaneidade*. Pontes Editores.
- Aranha, S. (2016). Os gêneros e as interações em teletandem institucional e integrado: Quais são, como são, o que são? *Revista Linguagem em Foco*, 8(1), 21–28.
- Aranha, S., & Leone, P. (2017). The development of DOTI (Data of oral teletandem interaction). In D. Fisher & M. Beibwenger (Eds.), *Investigating computer-mediated communication: Corpus-based approaches to language in the digital world*. Ljubljana: University of Ljubljana, Faculty of Arts (Vol. 1, pp. 172–192).
- Aranha, S., & Lopes, Q. B. (2019b). Moving from an internal databank to a sharable multimodal Corpus: The MulTeC case. *The ESPecialist*, 40(1), n.p.
- Brammerts, H. (1996). Tandem language learning via the internet and the International E-Mail Tandem Network. In D. Little & H. Brammerts (Eds.), *A Guide to Language Learning in Tandem via the Internet—CLCS occasional paper* (pp. 9–22). CLCS.
- Brammerts, H. (2003). Autonomous language learning tandem. In T. Lewis & L. Walker (Eds.), *Autonomous Language Learning in-Tandem*. Academy Electronic Press.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Canagarajah, S. (2020). Transnational work, translingual practices, and interactional sociolinguistics. *Journal of Sociolinguistics*, 24(5), 555–573. <https://doi.org/10.1111/josl.12440>
- Cappellini, M., Elstermann, A.-K., & Mompean, A. R. (2019). Reciprocity 2.0: How Reciprocity Is Mediated Through Different Formats of Learners' Logs. In C. Tardieu & C. Horgues (Eds.), *Redefining Tandem Language and Culture Learning in Higher Education*. Routledge.
- Celani, M. A. (2005). *ESP in Brazil: 25 years of evolution and reflection*. EDUC.
- Dooly, M. (2011). Divergent perceptions of telcollaborative language learning tasks: Task-as-workplan vs. Task-as-process. *Language Learning & Technology*, 15(2), 69–91.
- Elstermann, A.-K. (2017). *Learner support in telecollaboration* [Doctoralthesis]. Ruhr-Universität Bochum, Universitätsbibliothek.
- Kapac, P., & Schwienhorst, K. (2005). In two minds? Learner attitudes to bilingualism and the bilingual tandem analyser. *ReCALL*, 17(2), 254–268. <https://doi.org/10.1017/S0958344005000820>
- Lee, J. S., Hill-Bonnet, L., & Gillispie, J. (2008). Learning in Two Languages: Interactional Spaces for Becoming Bilingual Speakers. *International Journal of Bilingual Education and Bilingualism*, 11(1), 75–94. <https://doi.org/10.2167/beb412.0>

- Lewis, T. (2020). From tandem learning to e-tandem learning: How languages are learnt in tandem exchanges. In S. Gola, M. Pierrard, E. Tops, & D. Van Raemdonck (Eds.), *Enseigner et apprendre les langues au XXIe siècle*. Peter Lang B. <https://doi.org/10.3726/b16391>
- Picoli, F., & Salomão, A. C. B. (2020). O princípio da separação de línguas no teletandem: O que as teorias propõem e como ele funciona na prática. *Estudos Linguísticos (São Paulo. 1978)*, 49(3), 1605–1623. <https://doi.org/10.21165/el.v49i3.2458>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ramos, R. de C. G. (2008). ESP in Brazil: History, new trends and challenges. In M. Krzanowski (Ed.), *ESP and EAP in Developing and in Least Developing Countries* (Vol. 1, pp. 68–83). IATEFL; READCUBE. http://www.teachingenglish.org.uk/sites/teacheng/files/ESPBrazil_Ramos_.pdf.
- Rampazzo, L., & Aranha, S. (2018). A sessão oral de teletandem inicial: A estrutura retórica do gênero. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 34, 449–473. <https://doi.org/10.1590/0102-445063518072133051>
- Rampazzo, L., & Aranha, S. (2019). Telecollaboration and genres: A new perspective to understand language learning. *Journal of Virtual Exchange*, 2, 7–28. <https://doi.org/10.21827/jve.2.35637>
- Schwiehorst, K., & Borgia, A. (2006). Monitoring Bilingualism: Pedagogical Implications of the Bilingual Tandem Analyser. *CALICO Journal*, 23(2), 349–362.
- Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *JOSS*, 1(3). <https://doi.org/10.21105/joss.00037>
- Telles, J. A. (2015). Learning foreign languages in teletandem: Resources and strategies. *DELTA: Documentação de Estudos Em Lingüística Teórica e Aplicada*, 31, 603–632. <https://doi.org/10.1590/0102-4450226475643730772>
- Vassallo, M. L., & Telles, J. A. (2006). Foreign Language Learning in-tandem: Theoretical principals and research perspective. *The ESPecialist*, 27(1), 83–116.
- Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org/>.

Recebido em: 29/06/2023

Aceito em: 18/08/2023