DOI: 10.17469/O2107AISV000018

SIMONA SBRANNA, FRANCESCO CANGEMI, MARTINE GRICE

Quantifying L2 interactional competence

The Common European Framework of Reference for languages (CEFR) defines foreign language competence as communicative competence, emphasizing its interactional aspect. Nonetheless, L2 (second language) assessment often focuses on the quantification of grammar and lexical competence, neglecting interactional aspects, which are only subject to an impressionistic evaluation. With this exploratory study, we test a method for quantification and visualization of interaction management on L2 data. Our corpus includes 40 conversations both in L1 (first language) and L2 produced by Italian learners of German as L2. Results suggest that low levels of proficiency negatively affect the smoothness of the interactional flow, whereas the difference between L1 and L2 interactional patterns reduces with increasing L2 proficiency. Extracting reliable and testable metric, this method could represent a valid starting point to develop an instrument for a quantifiable assessment of interactional competence.

Keywords: L2 acquisition, communicative competence, interactional competence, assessment of L2 proficiency.

1. *Introduction*

The Common European Framework of Reference for languages describes learners as "social agents" (Council of Europe, 2001:9) who, as members of the society, constantly need to accomplish communicative tasks in various contexts and under different circumstances. To carry out the tasks they are faced with, they need to strategically combine their general and communicative language competence in the most appropriate way according to the context.

Figure 1 shows a schematic description of the communicative language competence extracted from the CEFR (Figueras, North, Takala, Van Avermaet, & Verhelst, 2009:32). Here communicative language competence is depicted as a compound competence formed by several components: linguistic competence, intended as lexical and grammatical knowledge, together with its cognitive organisation and accessibility; sociocultural competence, including the knowledge of register appropriateness, degree of formality, rules of politeness and the knowledge of linguistic rituals specific to a community; pragmatic and strategic competence, referring to the functional skills necessary to arrange the message and manage the conversation according to interactional schemata. The table also includes four different aspects of language activities learners can perform: production, reception, mediation and interaction.

Figure 1 - *Communicative Language Competence in the CEFR (Figueras et al., 2009:32)*

Table 4.2: CEFR Scales for Aspects of Communicative Language Competence

	RECEPTION		INTERACTION		PRODUCTION		MEDIATION	
	Listening	Reading	Spoken Interaction	Written Interaction	Spoken Production	Written Production	Spoken Mediation	Written Mediation
Linguistic Competence								
▪ General Linguistic Range	✓	✓	✓	✓	✓	✓	✓	✓
▪ Vocabulary Range	✓	✓	✓	✓	✓	✓	✓	✓
▪ Vocabulary Control			✓	✓	✓	✓	✓	✓
▪ Grammatical Accuracy			✓	✓	✓	✓	✓	✓
▪ Phonological Control			✓		✓		✓	
▪ Orthographic Control				✓		✓		✓
Socio-linguistic Competence								
▪ Socio-linguistic Appropriateness	✓	✓	✓	✓	✓	✓	✓	✓
Pragmatic Competence								
▪ Flexibility			✓	✓			✓	✓
▪ Turntaking			✓					
▪ Thematic Development	✓	✓		✓	✓	✓	✓	✓
▪ Cohesion and Coherence	✓	✓			✓	✓	✓	✓
▪ Spoken Fluency			✓		✓		✓	
▪ Propositional Precision	✓	✓			✓	✓	✓	✓
Strategic Competence								
▪ Identifying cues/inferring	✓	✓					✓	✓
▪ Turntaking (repeated)			✓					
▪ Cooperating			✓	✓				
▪ Asking for clarification			✓	✓				
▪ Planning					✓	✓		✓
▪ Compensating			✓	✓	✓	✓	✓	✓
▪ Monitoring and Repair			✓	✓	✓	✓	✓	✓

Production and reception are primary activities considered firstly in insolation since they are preliminary to mediation and interaction. Mediation consists in making communication possible among parties who are not able to directly communicate with each other, generally through the reformulation of an oral or written source message. Interaction is described as a complex process in which:

> at least two individuals participate in an oral and/or written exchange in which production and reception alternate and may in fact overlap in oral communication. Not only may two interlocutors be speaking and yet listening to each other simultaneously. Even where turn-taking is strictly respected, the listener is generally already forecasting the remainder of the speaker's message and preparing a response. Learning to interact thus involves more than learning to receive and to produce utterances (Council of Europe, 2001:14).

In its face-to-face form, interaction involves productive and receptive skills at the same time, as well as additional abilities which allow speakers to monitor the development of this process and constantly adjust to it in real-time.

As described into the CEFR, different abilities come into play at different moments of the interaction. In planning, speakers try to anticipate the forthcoming linguistic exchange, identify the common background information and opinion of the interlocutor to plan which conversational move best suits. In order to take the floor, speakers use turn-taking strategies and during the exchange, they co-operate

and build the conversation together. Speakers also need to be able to cope with unexpected misunderstandings and ask for clarification, or repair communication breakdowns, thus interlocutors constantly evaluate the on-going process to be able to appropriately react.

The description offered by the CEFR mirrors the definitions of interactional competence and interactional abilities provided by the existing literature. Jacoby and Ochs (1995) have described interaction as a form of co-construction, a joint creation of discourse between interlocutors. Hall (1993, 1995 as reported in He, Young, 1998) has argued that interactional competence is, therefore, context- specific since it emerges in varied interactive practices, to which participants contribute with the appropriate linguistic and pragmatic resources.

Moreover, especially in its oral form, interaction is the most common and direct way of communication in our daily life and for this reason, it should be ensured a central role in language teaching and testing. Indeed, the Framework states:

> High importance is generally attributed to interaction in language use and learning in view of its central role in communication (Council of Europe, 2001:14).

However, assessment of learners' proficiency often tends to focus mainly on grammar and lexicon, neglecting or even excluding interactional aspects. Many language test formats only involve a written form, such as the cloze format, where some words of a text are replaced with gaps to be filled in by learners. These kinds of test are often used with the explanation that they show correlations to all receptive and productive abilities (reading, listening, writing and speaking). Nevertheless, they mainly put to test grammar and vocabulary knowledge, leaving out the full range of pragmatical and strategical resources required in oral interaction.

Bachmann and Palmer in "Language Testing Practice" (1996) mention interactiveness as one of the fundamental characteristics a good quality language proficiency test should have: reliability, construct validity, authenticity and interactiveness. Reliability refers to the consistency of the measurement, thereby of the results given by the test. Construct validity indicates the possibility to interpret the score of the test as a valid indication of global language proficiency. Authenticity defines how correspondent the task given to learners in test circumstances is to real-life tasks they would perform using the L2. Finally, interactiveness refers to the degree of involvement of learners' different abilities in accomplishing the task, i.e. the extent to which a test involves various learners' skills, which include general language knowledge, metacognitive strategies and strategic competence for planning and dealing with unexpected difficulties, topical knowledge and affective schemata, which refers to learners' emotional response to the task. According to the authors, tasks with a high level of interactiveness are role play and long conversations, as they require learners to draw on all these abilities.

Some possible reasons for neglecting highly interactive tasks in language proficiency testing may be practical. He and Young (1998) point out that having learners interviewed by native or highly proficient speakers can create certain difficulties. First, such interviewers have to be available; secondly, the interviews need to be

carried out for a reasonable length of time to allow the interviewer to elicit enough linguistic data from the learner so that these data can be considered representative of the learner's global knowledge. Hence, such testing would require more assessors at the same time, be time-consuming and consequently more expensive than a test format, such as the cloze test, which can optimize time for testing and correction. Furthermore, the quantification of the skills involved in interaction may turn out to be extremely time-consuming and complex to synthesise. As a result, interactional competence is often subject only to a qualitative evaluation based on illustrative scales[1], the interpretation of which may include a certain degree of subjectivity.

Intending to propose quantification methods for L2 speaking ability, research has mainly focussed on the measurement of fluency. Indeed, fluency is consistently mentioned as a fundamental component of learners' oral performance in various assessment traditions and its correlation with general L2 proficiency has been demonstrated by several studies (De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; De Jong, Groenhout, Schoonen & Hulstijn, 2015; Segalowitz, Freed, 2004).

In the following paragraphs we will shortly review theories and findings on fluency as a measure of L2 speaking proficiency.

2. *Background*

Fluency has been identified as one of the main aspects ensuring the success of the speaking performance (De Jong, 2016). One of its first definitions can be traced back to Fillmore (1979). He defines fluency as a measure of how well a language is spoken, in other words the skill to use L2 knowledge efficiently, and enumerates four dimensions of fluency including both quantitative and qualitative aspects: the ability to speak at length with few breaks; the ability to speak in a coherent, reasoned, and semantically dense way; the ability to talk appropriately according to the context; the ability to be creative and imaginative in speech production.

Starkweather (1987) suggests instead four dimensions of fluency mainly related to physical aspects of speech: continuity, rate, rhythm, effort. In other words, fluent speech should present few discontinuities, have a regular rhythm and a fast rate, and not require too much cognitive and physical effort (Zmarich, 2017).

In his model of fluency, Logan (2015) adds to the described eight dimensions two additional ones: naturalness, i.e. how much speech resembles that uttered by a typical speaker with regard to continuity, rate, rhythm and effort; and stability, i.e. how similar are speaker's performances over time if subject to repeated measurements.

Segalowitz (2010) focuses on L2 fluency from a dynamical system perspective. He argues that fluency is strongly linked to the social context in which the speech

[1] An example from the illustrative scale for overall spoken interaction (C1 level): "Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language" (Council of Europe, 2001:74).

performance takes place and distinguishes three aspects: utterance fluency, cognitive fluency and perceived fluency.

L2 utterance fluency refers to the fluidity observable in a speech sample and quantifiable by temporal measures, among which the author mentions syllable rate, duration and rate of hesitations, filled and silent pauses, breakdown fluency (pausing phenomena) and repair fluency (false starts, corrections, repetitions). Indeed, most studies calculating fluency temporal measures follow the classification in the sub-components breakdown, repair and speed fluency (Tavakkoli and Skehan, 2005).

L2 cognitive fluency refers to the fluidity of the cognitive processes underlying speech production, such as processing skills (declarative and procedural knowledge), efficiency and speed of semantic retrieval, and cognitive load in working memory. Some measures of cognitive fluency have been found to correlate with L2 proficiency, e.g. reaction time and switch cost measures. Reaction time speed and its coefficient of variability have been used to operationalise the efficiency of semantic retrieval (Segalowitz, Freed, 2004), while switch cost measures has been used as an indicator of linguistic attention, which refers to attention shifting guided by connections among grammatical elements within utterances (Duncan, Segalowitz & Phillips, 2014).

Such systemic understanding of fluency is assumed also by Kormos' psycholinguistic model (2006) in which different cognitive processes underlie the three above-mentioned sub-components of fluency. In particular, breakdown measures are related to learners' effort, such as final-clause pauses, which reflect learners' conceptualization and planning of the message, and mid-clause pauses, which represent the time taken by learners to encode and formulate the linguistic information; repair measures signal the monitoring of the speech output and consequently the amount of attention required for speaking in L2; and finally, speed-related measures inform about the degree of automatization of all these processes.

Another important aspect to take into account is the fact that both utterance and cognitive fluency are specific to each person. Still, individual variability in L1 can only partially explain individual variability in L2 (De Jong *et al.*, 2013) since disfluency is also characterised by L2-specific features, e.g. a higher cognitive load. Therefore, it may be a good scientific practice to consider L1 fluency measures as a baseline for each individual's fluency characteristics (as in De Jong *et al.*, 2015; Saito *et al.*, 2019) to get a clearer picture of L2-specific fluency measures by partialling out the variables that are not especially related to L2 disfluency phenomena (Segalowitz, 2010).

Finally, L2 perceived fluency indicates subjective listeners' ratings on how fluent a speaker is. One disadvantage is that being subjective, perceived fluency is only moderately informative about utterance fluency and cannot explain all the variance of its objective measures. However, it is helpful to get an understanding of what cues are relevant to native listeners when judging L2 speech fluency in relation to L2 proficiency. Moreover, a listener's judgment of their interlocutor's fluency can affect the interaction and influence both speakers' fluency.

2.1 Fluency measures and operationalisations

Research on L2 fluency has focused on individuating which objective measures can better explain L2 fluency judgments and has mainly concentrated on temporal features.

A categorisation of aspects of fluency comparable to the more recent triad "breakdown, repair and speed" (Tavakkoli, Skehan, 2005) was already proposed in one pioneering study. Riggenbach (1991) classifies the features which can characterise a judgement of fluent or non-fluent in non-native speech into hesitation and repair phenomena, and rate and amount of speech. The study also includes an analysis of interactive features contributing to the turn-taking alternation, such as overlaps, pauses between turns and collaborative completitions. Hesitation phenomena and speech rate were found to be significantly correlated to ratings of L2 fluency, with hesitations placement and the resulting discourse chunking playing a central role. On the contrary, results related to repair phenomena appeared to be less clear, probably due to the small set of data. The same holds for interactive features, which revealed a high variability because of the idiosyncratic nature of interactions that vary according to many linguistic and non-linguistic factors. Furthermore, this pragmatic- oriented analysis is referred to be extremely time-consuming and, indeed, studies on turn-taking fluency are relatively scarce.

Later on, many studies confirmed that perceived fluency by both native listeners (Kormos, Dénes, 2004; Derwing, Rossiter, Munro, & Thomson, 2004; Rossiter, 2009; Bosker, Pinget, Quené, Sanders & De Jong, 2013; Préfontaine, Kormos, & Johnson, 2016; Saito, Ilkan, Magne, Tran, & Suzuki, 2018; Suzuki, Kormos; 2019) and L2 listeners (Magne, Suzuki, Suzukida, Ilkan, Tran, & Saito, 2019) is closely related to the speed of delivery and pausing phenomena.

As reported in Suzuki and Kormos (2019), a more recent approach in research on fluency differentiates three independent dimensions of breakdown fluency – frequency, duration and location of pauses – and all of them have been demonstrated to independently contribute to fluency. Moreover, concerning pause location, mid-clause pauses have been found to have a more distinctive role than final-clause pauses (Saito *et al.*, 2018; Magne *et al.*, 2019; Suzuki, Kormos; 2019). One possible reason, following Kormos' model (2006), would be that mid-clause pauses, being associated with the time required for linguistic encoding, are more representative of proficiency than final-clause pauses associated with content planning.

However, these studies present several differences in methodology, in particular regarding sample size, rating methods, task used for data collection and operalisation of measures. The following table summarises the most listed temporal measures in literature reviews on fluency and some interactional measures.

Table 1 - *Most used temporal and interactional measures in research on fluency*[2]

Speed of speech measures	**Formula / Description**	**Reference**
speech rate	n° of syllables / total time	Kormos (2006) in: De Jong (2016)*
pruned speech rate	n° of syllables - n° of disfluent syllables / total time	De Jong (2016)
phonation time ratio	speaking time / total time	Kormos (2006) in: De Jong (2016)*
articulation rate	n° of syllable / speaking time	Kormos (2006) in: De Jong (2016)
mean lenght of syllables	speaking time / n° of syllable	De Jong (2016)*
mean lenght of run	n° of silent pauses / n° of syllables	Kormos (2006)*
Breakdown fluency measures		
Frequency		
number of pauses	n° of pauses / total time or speaking time	as in Saito et al. (2019)*
number of silent pauses	n° of silent pauses / total time or speaking time	Kormos (2006) in: De Jong (2016)*
mean lenght of utterance	total speaking time / n° of utterances	De Jong (2016)
numer of filled pauses	n° of filled pauses / total time or speaking time	Kormos (2006) in: De Jong (2016)
Duration		
duration of silent pauses	pausing time / n° of silent pauses	Kormos (2006) in: De Jong (2016)
Location		
mid-clause pauses	interrupting the clause	Saito et al. (2019)*
final-clause pauses	juncture pauses (clause and phrase boundaries)	Saito et al. (2019)*
Repair fluency measures		
n° of repetitions	n° of repetition / total time or speaking time	De Jong (2016)*
n° of repairs	n° of corrections and restarts / total time or speaking time	De Jong (2016)*
Turn-taking fluency measures		
pause	within speaker silence	Heldner & Edlund (2010)
gap	between speakers silence	Heldner & Edlund (2010)
overlap	turn-changing and non-turn-changing	Heldner & Edlund (2010)
backchannel	not taking the turn	Riggenbach (1991)
collaborative completion	attempt to complete a sentence or a phrase of the other	Riggenbach (1991)

Moreover, tasks used in experimental and assessment settings can have a large impact on L2 fluency. Derwing *et al.* (2004) argue that the task used to collect data can exert an influence on learners' fluency depending on its degree of freedom, since picture narrative and description impose a given range of lexicon and syntactic structures, while a monologue, or a conversation with free choice of topic allow learners to have much more control on the content and the expressions used to deliver it.

In non-experimental environments, assessment of L2 oral proficiency is often conducted through oral proficiency interviews, where a native speaker tries to elicit linguistic information from learners using a script representative of real-life language use settings. However, although these interviews try to simulate ordinary conversation, they present various constraints that can affect fluency (He, Young, 1998), e.g. interviews take place in an institutional setting; speech activities are predetermined; participants have different statuses, two different L1s and cultural backgrounds, as well as two different proficiency levels of the language used during the interview. As there are numerous factors which can negatively affect learners' fluency performance, a wider range of speaking tasks, in particular more open and interactive ones, should be employed.

[2] For the references of the studies in which the measures signalled with a star have been found to be significant see Saito *et al.* (2019).

For these reasons, especially in experimental settings, assuming L1 fluency measures as a baseline for each speakers' L2 fluency can help to explain some idiosyncratic differences by controlling for non-linguistic factors possibly affecting learners' performance (Segalowitz, 2016), such as contextual factors (e.g. attitude to the task and the interlocutor), but also learner-specific ones (e.g. personality and motivation).

2.2 Turn-taking fluency visualisation methods

Most studies mentioned in the previous paragraphs have focussed mainly on the concept of fluency as an individual phenomenon. Their main concern was to define if and how much a speaker is fluent, its perception from listeners' perspective and implications for L2 assessment, considering also that in many experiments monologic tasks were used for data collection.

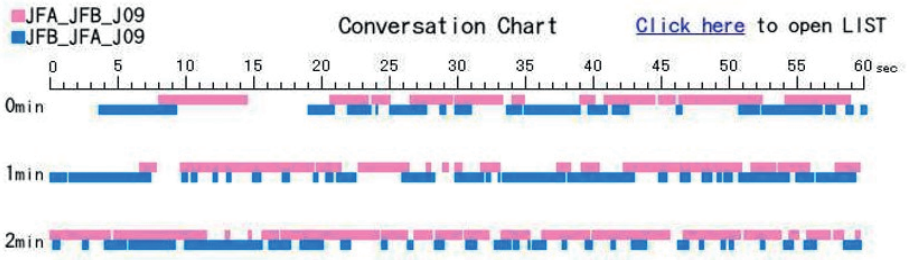
However, researchers agree that the circumstances in which learners' performance takes place is fundamental to make assumptions about the performance itself. Therefore, the measures above proposed only provide incomplete information about L2 fluency if considered out of context. For example, it has been suggested that speed of delivery and pausing behaviour are accommodated to the interlocutor during the interaction (Kourisidis, Dorran, 2009) so that a jointly achieved harmonisation of tempo occurs. This phenomenon has been depicted through the metaphor of interactional "flow" (McCarthy, 2009).

Moreover, being ruled by the turn-taking system (Sacks, Schegloff, & Jefferson, 1974), dyadic interactions better put to test learners' degree of automaticity. Indeed, turn-boundaries (also called transition relevant places – TRP – in the field of conversational analysis) are the places in which smooth or disfluent transition of turns can take place and which require to appropriately anticipate the end of the interlocutor's turn to be able to quickly react (Levinson, 2015; Bögels, Torreira, 2015).

For these reasons, judgements of fluency based on a single speaker and ignoring the interlocutors' contribution to the conversation would lack the interactive perspective and miss important information about learners' ability to co-create fluency (remember that interaction is also described as a co-creation process in the CEFR).

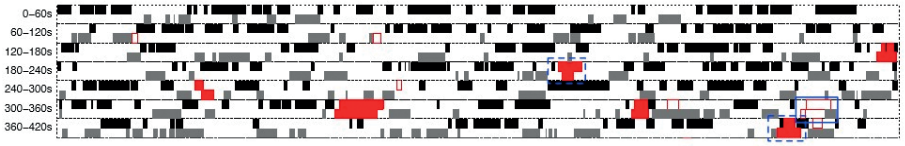
Some studies focussing on conversational speech rhythm have developed visualisation tools to represent the speech activity performed in dyadic exchanges. In his study on speech activity patterns in telephone conversations, Campbell (2007) created a plot to visualise the timing organisation of the interaction using four classes of activities: both interlocutors silent, both interlocutors talking simultaneously, only interlocutor A speaking, only interlocutor B speaking (Fig. 3). On the horizontal axis of the plot, a time window of one minute is displayed, whereas the vertical axis reproduces time passing throughout the interaction. The speaking activity was then represented on this graphic scaffold by colour-coded bars, whose length represents the duration of each speakers' turn. As a consequence, when the two different bars are upon each other, speakers are speaking at the same time causing an overlap, whilst when bars interrupt and a white space follows, speakers are being silent.

Figure 2 - *A section of Campbell's conversation chart (2007:344)*



The same type of plot has been used to display laughter in dyadic conversations (Trouvain, Truong, 2013) to investigate the interplay of laughing and speaking activities (Fig. 3). In this plot the four classes of activities are of different nature: filled and empty red bars signal respectively overlapping and non-overlapping laughs, whereas filled dark and grey bars represent speakers' speech (blue rectangles mark instead cases of interest to the study).

Figure 3 - *Visualisation of speech and laugh activity in dyadic conversation (Trouvain et al., 2013:2)*



This kind of charts displaying the duration and timing of specific classes of interest has two main benefits. It serves as an "eye-opener" (Trouvain, Truong, 2013:4) helping researchers to evaluate their intuitions by means of a visual exploration and comparison of data. Additionally, the data on duration and timing extracted to create the plot provide the material for acoustic analysis and statistics for hypothesis testing.

In this study, we test this type of visualisation plot on L2 interactional data to assess its representative power and usefulness in capturing differences in oral interaction management by learners with different levels of L2 proficiency. Due to its preliminary nature, we do not conduct any hypothesis testing, but instead concentrate on the application of the method, i.e. the extraction of metrics, the quantification of L1 and L2 interactions and the discussion of analysis results. To fill the gap in quantification methods for interactional competence, we suggest these visualisation and quantification tools as possible groundwork for developing a more complete instrument for a standardised assessment of L2 interactional competence.

3. *Method*

As a proxy for interactional competence, we explore the degree of fluency of the interaction co-created by participants through the turn-taking system. Following Campbell (2007), we operationalise the interactional flow by quantifying four classes of activities: the percentage of time each speaker takes the floor (1, 2), how long the speech of the participants overlaps (3) and the amount of total silence (4).

Subjects are 40 Italian students of German as L2 with different L2 proficiency levels assessed by the German language courses they were attending. Speakers were aged between 19 and 65, 33 were female and 7 were male[3]. Some learners were students at the "Goethe Institut" of Naples and others at the "Università degli Studi di Napoli L'Orientale", with German as foreign language as one of the main subjects[4]. Speakers were matched according to their L2 proficiency level to perform the recording in pairs. In table 2 the dyads of students who performed the task together and their corresponding level of German are listed. However, because of constraints due to subjects' time availability, dyads from 11 to 14 were mixed and are thus composed by an interlocutor with B1 level and the other with B2 level.

Table 2 - *Subjects matched to perform the task for data collection according to their level of L2 German. Dyads from 11 to 14 are mixed and composed by interlocutors with B1 and B2 levels respectively*

DYAD	LEVEL	DYAD	LEVEL
1	A1	11	B1-B2
2	A2	12	B1-B2
3	A2	13	B1-B2
4	B1	14	B1-B2
5	B1	15	B2
6	B1	16	B2
7	B1	17	B2
8	B1	18	B2
9	B1	19	C1
10	B1	20	C1

The corpus consists of forty dialogues, twenty in the native language – Italian, Neapolitan variety – and twenty in German as L2, following the suggestion to use

[3] Apart from one student of the Goethe Institut aged 65, all other participants were aged between 19 and 38. Median age of participants = 21; standard deviation = 7,94.

[4] 24 learners had benefited from a stay in German-speaking countries for a variable length of time (from one to ten months) either for a short language course, or an exchange period at a partner university. However, the effect of a period abroad is neither accurately quantifiable, nor equal for everyone. It varies according to the amount of input and use of the foreign language (consider, for example, exchange students who do not manage to establish regular contact with the locals, or decide not to attend a German language course). That said, since a period of immersion in the foreign language contributes to language proficiency, we did not consider this variable separately.

learners' L1 as a baseline against which to assess their L2 interactional patterns (Segalowitz, 2016). Indeed, we supposed that their way of interacting would not be similar to a native German one, since our subjects are living in Italy, studying in Italy and currently talking to an Italian interlocutor with whom they share the same L1 and culture. There is no foreign exposure which could favour them in approaching a German native style. Moreover, using the L1 as a baseline was informative in a previous exploration phase (Sbranna, Cangemi, Grice, 2019) conducted on two dyads, one with beginners and the other with advanced learners. The pilot study showed that with a higher L2 proficiency the interaction in L2 approached the same interactional pattern learners used in their own L1. Recordings were made at the "Goethe Institut" in Naples using headset microphones (AKG C 544 L) connected through an audio interface (Alesis iO2 Express) to a computer running *Praat* (Boersma, 2001).

The spontaneous speech data were elicited using the Map Task (Anderson *et al.*, 1991; Grice, Savino, 2003 for set up, map layout and instructions), a task-oriented dialogue which matches a kind of task described in the CEFR, the goal-oriented co-operation task. Participants sit opposite each other and have no eye-contact. They are provided with two maps; one speaker receives a map with a route drawn across landmarks – instruction giver – and has to describe the route to the other participant – instruction follower –, whose map only has landmarks. The goal is to co-operate so that the instruction follower can reproduce the route on their map thanks to the instructions given by the partner. Some landmarks are however different and speakers only discover it during the task, which creates unexpected problem-solving situations.

This task was chosen for two reasons. It can be performed at every proficiency level since learners should address the topic of grammar and vocabulary knowledge related to road indications at a beginner level according to the CEFR. In addition, following the suggestions of previous studies, it presents a fair degree of openness thanks to the unforeseen unmatched landmarks increasing the degree of spontaneity in interaction.

Participants first read an Italian version of the game instructions and carried out the task in their native language. Afterwards, before performing the task in L2, they watched a video with a German native speaker explaining the instructions again in German to help them get into the language and reduce the L1 bias. They kept the same role (either instruction giver, or follower) in both languages to prevent cross-language differences in their fluency from being attributable to the factors regarding their role in the task.

At the end of the recording session, learners were provided with an online test for lexical competence, the German version of Lextale (Lemhöfer, Broersma, 2012), originally designed to test English as L2 in an experimental setting and found to be a reliable indicator of general L2 proficiency.

3.1 Procedure

The annotation and extraction procedure is composed of three steps. After having extracted the two channels from the stereo recordings, a first step consisted in the

automatic labelling of interpausal units in *Praat* using the function of silent interval detection. Secondly, boundaries were manually checked and corrected to make sure that interpausal units were rightly identified, since some voiceless consonants were automatically labelled as silence. Finally, the Text Grids resulting from the two channels of each dialogue were used as input files for a *Praat* script, which generated a figure depicting each speaker's contribution to the interaction as it develops over time (Fig. 4, 5, 6, 7).

Each horizontal bar corresponds to an interpausal unit uttered by one of the two speakers involved in the task. Speakers are colour-coded with red being the instruction giver and black the instruction follower. Time unfolds from top to bottom – minutes –, and from left to right – seconds – so that the interaction can be followed as on a written page in a left-to-right writing system.

Figure 4 and 5 display a low proficiency dyad performing the task in L1 and L2. The first striking difference is the total length; in L2 these speakers need roughly half of the time more than in L1 to conclude the task. Moreover, differently from the L1 smooth pattern, the flow of interaction in L2 appears much more fragmented, with shorter turns and more frequent and longer pauses, especially during the turn of an individual speaker.

This observation is in line with the systemic perspective of fluency mentioned in the background, according to which utterance fluency measures mirror learners' cognitive fluency. Indeed, in the case of this dyad, who has a low proficiency in German (A1 level, beginner), it is not surprising to find lengthened within- speaker pauses in L2 as compared to their interactional behaviour in L1, since this reflects a high cognitive effort necessary for the retrieval of linguistic information and formulation of the linguistic message they want to convey to their interlocutor.

Figure 4 - *Visualisation of interactional flow (L1 Italian by dyad 1 – low L2 proficiency)*

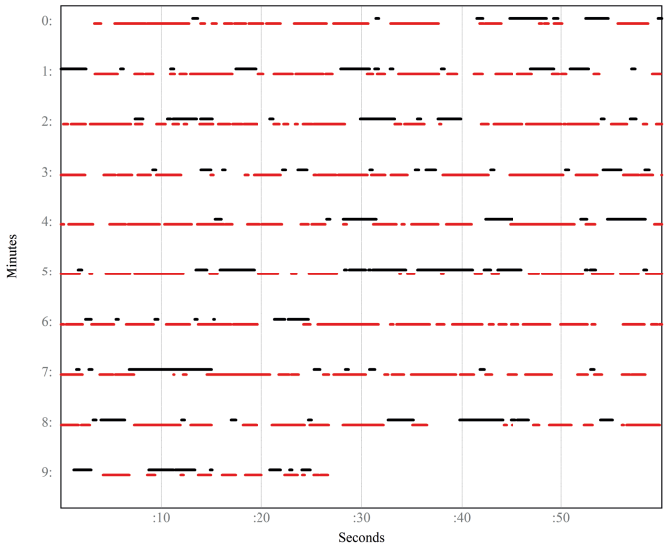
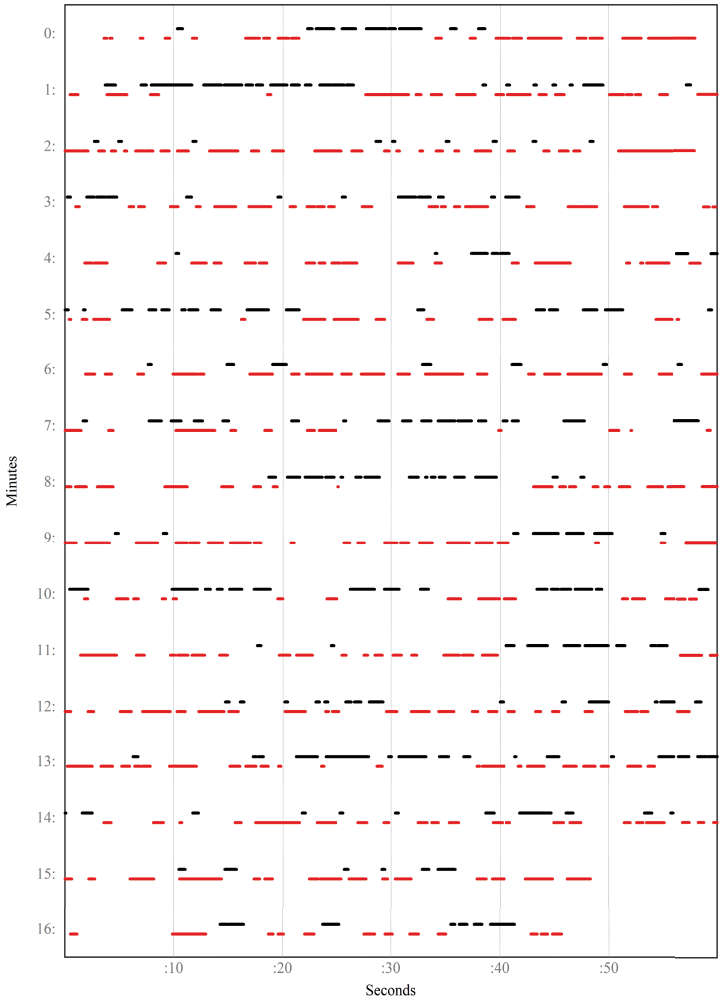


Figure 5 - *Visualisation of interactional flow (L2 German by dyad 1 – low L2 proficiency)*



Figures 6 and 7 show the interactional pattern of a dyad with high proficiency in L2 German (C1 level, advanced). In this case, it is difficult to identify at first sight which dialogue was carried out in the foreign language since the two interactional patterns look very similar. We can still notice a that in L2 the two highly proficient speakers need a little more time to complete the task, yet the difference is extremely slight and may be due to other factors generating variability in total duration of the interaction.

Figure 6 - *Visualisation of interactional flow (L1 Italian by dyad 19 – low L2 proficiency)*

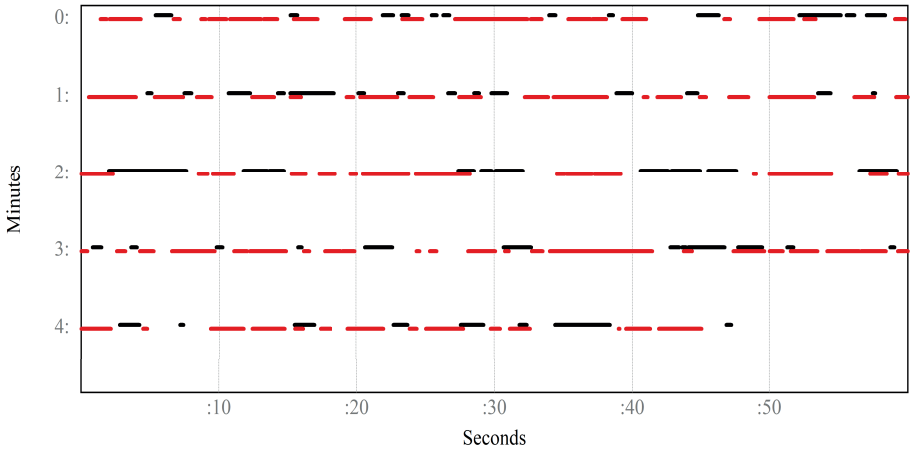
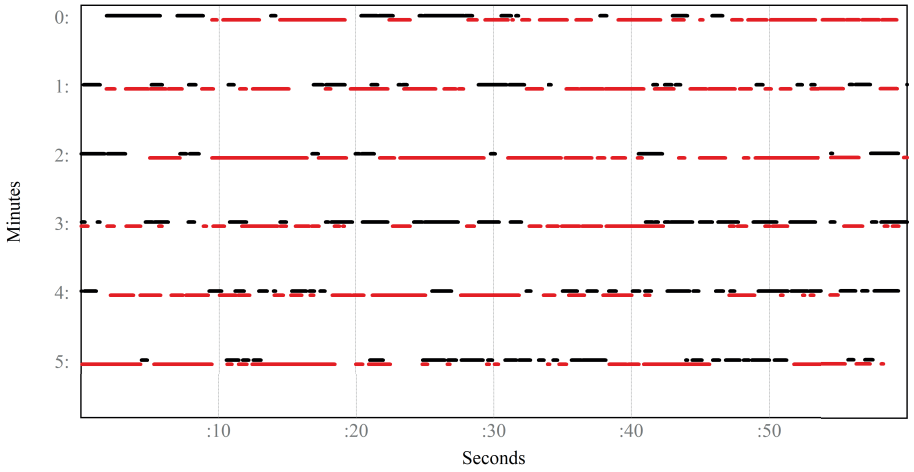


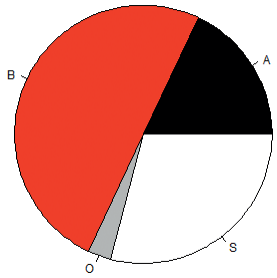
Figure 7 - *Visualisation of interactional flow (L2 German by dyad 19 – low L2 proficiency)*



The first exploration of data using these visualisation plots suggests that a higher proficiency level in L2, thanks to an enhanced automatization of the cognitive processes required to speak a foreign language, enables a degree of smoothness in managing the interactional flow closer to the one learners have in their native language.

In addition to this figure, the *Praat* script derives from the extracted data a table used to generate pie plots (example in Fig. 8) in *R* (R Core Team, 2013). The four sections of the pie plots use the same colour-coding of the visualisation plot to show the percentages of speech uttered by each speaker (A, instruction follower in black; B, instruction giver in red), the total amount of silence (S, in white) and overlap between speakers (O, in grey). The radius of the circle represents the total duration of the interaction, so that the bigger the pie is, the longer the interaction.

Figure 8 - *Example of pie plot summarising the interactional flow*

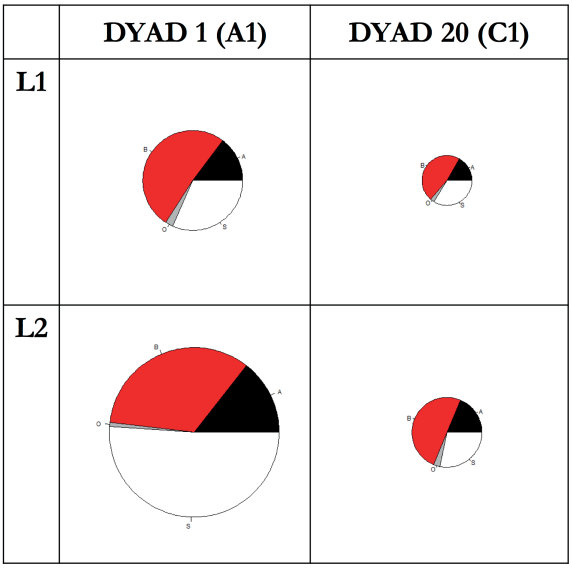


While the visualisation plot is useful to observe the time-aligned development of the interaction, this pie plot is helpful to summarize and quantify the partition of the interactional pattern into the four classes of activities and its total duration.

4. *Results*

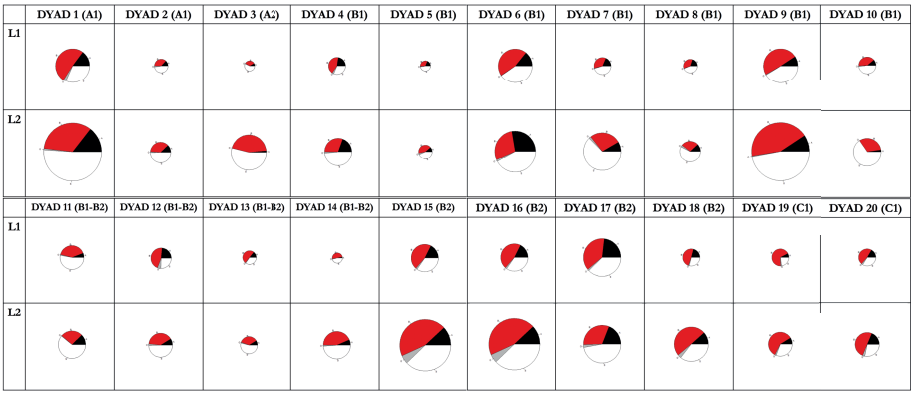
The pie plots described above capture the changing of the interactional patterns with increasing command of L2. As an example of their informativeness, we show four pie plots representing four dialogues performed by two dyads in L1 and L2 (Fig. 9): on the right there are two learners with low L2 proficiency and on the left two learners with high L2 proficiency.

Figure 9 - *Pie plots for dyad 1 (low proficiency – A1, on the left) and dyad 20 (high proficiency – C1, on the right). Plots in the upper line display dialogues in L1 and plots below dialogues in L2*



The dyad with high proficiency of German (C1) presents two very similar patterns of interaction across languages. The ratio of time speaking between the giver and the follower remains 3:1 when they repeat the task in L2. Differently, the dyad with low proficiency of German (A1) features two very different interactional patterns. In L1 the ratio of time speaking between the giver and the follower is 3:1, whereas in L2 the ratio changes to 2:1, with the giver speaking less in L2 than in L1. Further, more than half of the conversation is in silence. The following figure depicts the pie plots for all interactions of the corpus (Fig. 10).

Figure 10 - *Pie plots for the whole corpus ordered by increasing proficiency level of L2 German*

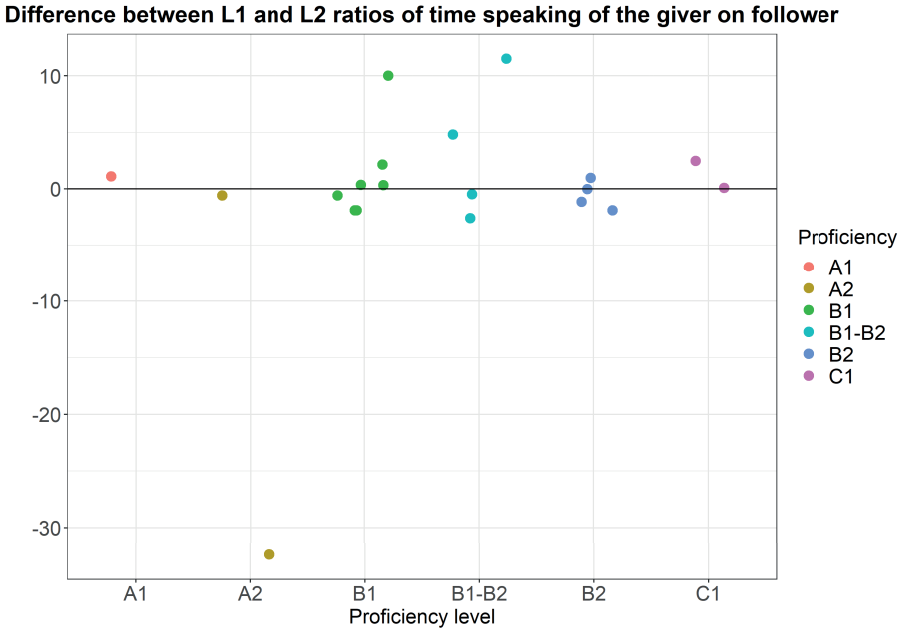


One commonality is that all interactions in L2 have a longer duration compared to those produced by the same dyad in L1. However, L2 interactional patterns start consistently resembling those in L1 from the dyads at B2 level of competence. Below this level interactional patterns in L1 and L2 reveal high variability across the two languages (except for dyads 2, 6 and 9).

To summarize data contained in all pie plots, we derived a graph calculating the difference between L1 and L2 ratios of time speaking between the giver and follower (Fig.11), so that each point in the graph represent a dyad and not a single speaker for the above-stated reason that interaction performances should not be analysed or evaluated in isolation as they are a form of co-creation of discourse by both – or, in case of multi-party conversations, all – interlocutors.

As in a first exploration, we noticed that the difference between the L1 and L2 ratios of time speaking between the giver and the follower is around 0 for high proficient learners (B2 and C1) and different from 0 for less proficient learners, we expect that with increasing proficiency, points will approach to 0 as indicating less difference in the interactional behaviour learners have across languages.

Figure 11 - *Difference between L1 and L2 ratios of time speaking of the giver on follower. The x-axis displays the proficiency level of L2 German, the y-axis shows the values resulting from the formula. The more points approach to 0, the less difference there is between learners' interactional behaviour in L1 and L2*



The graph shows high idiosyncratic variability. To start with the beginner group (A1 and A2 levels of proficiency), we can notice that two dyads are near to 0 and, therefore, present few differences in learners' interactional behaviour between L1 and L2. Indeed, grammatical and lexical resources are not the only factor contributing to conversational rhythm, since many other linguistic and extralinguistic factors play a role, e.g. personality, engagement in the task, relationship between speakers, not to forget the skill to strategically draw on the few resources beginners have to reach the goal of the interaction.

The golden outlier represents a dyad with particular behaviour – the follower only utters a few sentences towards the end of the dialogues in both L1 and L2 –. Especially in this case, having used the L1 as a baseline for learners' interactional behaviour has revealed to be useful, because this piece of data would have generally been associated with poor command of the L2, while we could observe that this speaker behaves exactly the same in L1, possibly due to lack of motivation and engagement in the task. Nevertheless, not disposing of many samples for the beginner level of proficiency, these dyads cannot be considered informative about a general trend.

On the other hand, the high variability displayed in the intermediate groups (B1 and mixed B1-B2 levels of proficiency) tends to reduce in the more advanced B2 and C1 levels. This observation would confirm the hypothesis emerged from the exploratory phase, i.e. learners with a higher proficiency of the L2 tend to repro-

duce in L2 the same interactional pattern they produce in their native language. In other words, less cognitive load, time for information retrieval and formulation, and attention required in L2 due to a higher degree of automatization allow learners to gain in smoothness of the interactional flow. Nonetheless, having only two dyads for the C1 level, a more conspicuous and homogenously distributed number of samples across proficiency groups would be required to confidently test this pattern.

Finally, to check an eventual relation of total duration of the dialogue and of lexical competence with the interactional pattern, we created two additional graphs and reduced the range of values on the y-axis cutting out the golden outlier to better visualise the remaining points (Fig. 12, 13).

In figure 12, the dimension of the circles represents the score obtained in the test for vocabulary knowledge by the giver, who is generally the one leading the task, so that the larger the circle is, the higher is the score they received. In figure 13, the dimension of the circles results from the difference in dialogue duration between L1 and L2, so that the bigger the point is, the longer is the time learners needed to accomplish the task in L2.

Figure 12 - *Relation between lexical competence and interactional pattern. The x-axis displays the proficiency level of L2 German, the y-axis shows the difference between L1 and L2 ratios of time speaking of the giver on follower. The dimension of the circles represents the score for givers' vocabulary knowledge*

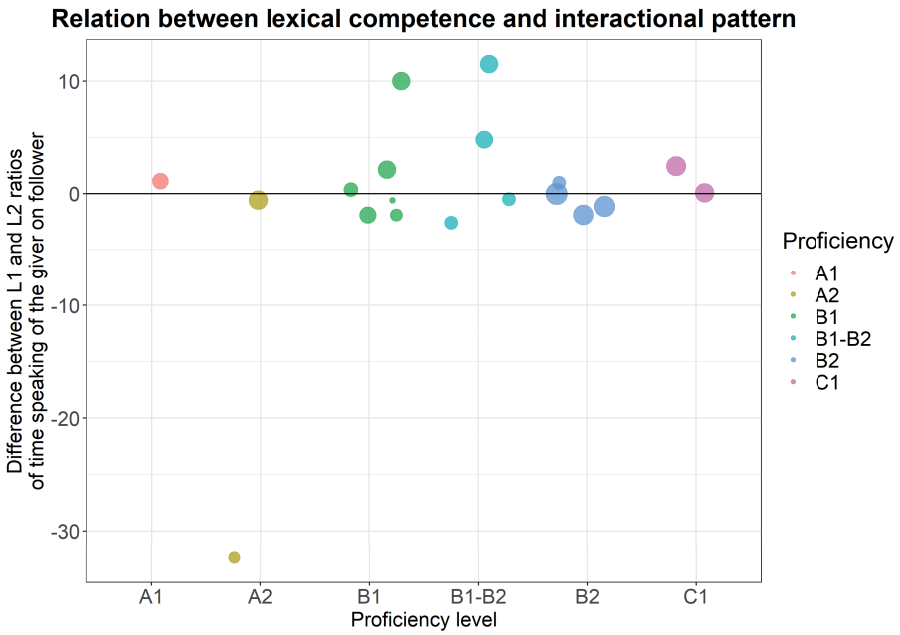
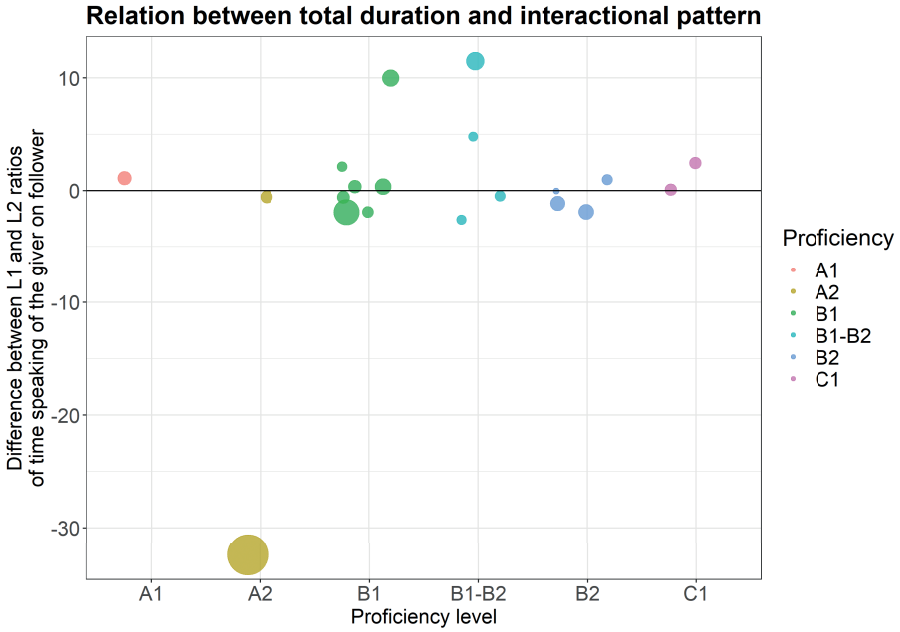


Figure 13 - *Relation between total duration of the interaction and interactional pattern. The x-axis displays the proficiency level of L2 German, the y-axis shows the difference between L1 and L2 ratios of time speaking of the giver on follower. The dimension of the circles represents the difference in dialogue duration between L1 and L2*



In neither case, it is possible to observe a clear trend. Especially in the graph displaying vocabulary scores (Fig. 12), in B1, B1-B2 and B2 groups both very low and high scores are near to zero, suggesting that lexical competence does not seem to be a factor determining how different the interactional patterns are in L1 and L2. Neither the second graph (Fig. 13) shows a clear pattern, but as the total duration of interaction may be at least partly dependent on a higher number of breakdown phenomena and slower speed of speech, which are directly related to L2 proficiency, it could be worth testing the relation between the difference in dialogue duration between L1 and L2 and interactional patterns on a larger corpus, which might provide a better understanding of the relationship between these two factors.

5. *Conclusion*

In this contribution, we problematised the absence of a standardised instrument for the quantification of interactional competence in L2. To open up new perspectives for L2 assessment, we focussed on the development of visualisation tools and a quantification method that can extract reliable and testable metrics for interactional aspects of communication.

We tested its informativeness on a corpus of L1 and L2 interactions including different levels of L2 proficiency. We found that more proficient learners are able

to maintain the natural interactional rhythm they have in L1 in their L2, as demonstrated by the interactional patterns displayed. We also observed that lexical competence does not seem to influence learners' interactional behaviour, which would suggest that mastering the lexicon does not automatically ensure a higher degree of success in oral interactions. Moreover, lexical scores do not seem to lead to corresponding levels of general competence. Indeed, the learning process is not linear and there is no discrete order in L2 knowledge acquisition (Nava, 2010), so that different skills can improve at different speeds. Since open interactional tasks put to test learners' L2 abilities in a more comprehensive way, an enhancement of interactional tasks in L2 experimental and testing settings can help to obtain a clearer picture of learners' L2 general proficiency and possibly shed light on the interplay among the different skills.

However, our corpus only included a few samples for beginner and advanced groups, in contrast to the more conspicuous intermediate group. Using the proposed metrics to draw statistical inferences on a corpus with a homogeneous number of samples across all L2 proficiency levels would make it possible to test these observations and collect additional evidence.

Despite this limitation in the sample, the method proposed was shown to be highly beneficial for the exploration and analysis of L2 oral interactions. It permits an immediate comparison of learners' interactional behaviour in L1 with their performance in L2 and can be used to observe the developmental trajectory of L2 interactional competence across different stages of the learning process. For these reasons, this method of visualisation and quantification of oral interactions could represent a starting point for quantifying L2 interactional competence in a standardised way.

In this study, we based our analysis on temporal measures only and did not provide an analysis of the transitions, i.e. sequences in which speakers change turns. A future development of this tool contributing to a more fine-grained analysis would be the integration of measurements for pragmatic and strategic competence from the CEFR scale for interaction (e.g. asking for clarification, compensating, cooperating, monitoring and repair). One option may be an analysis of conversational moves to assess the strategies learners use to coordinate the interaction throughout the learning process. However, such analysis presents some disadvantages inasmuch as it is time-consuming and may provide extremely variable results due to the highly idiosyncratic nature of interaction. Therefore, a pilot study evaluating the time cost and the informativeness of different pragmatic measurements about L2 interactional competence needs to be carried out to identify which ones can best and most effectively contribute to a fully comprehensive instrument for L2 assessment purposes.

Acknowledgement

We would especially like to thank the Goethe Institut of Naples for generously providing rooms for the recording sessions. Financial support for carrying out the

research is gratefully acknowledged from a.r.t.e.s Graduate School and CRC-1252 "Prominence in Language", funded by the German Research Foundation.

Bibliography

Anderson, A.H., Bader, M., Bard, E.G., Boyle, E.H., Doherty, G.M., Garrod, S.C., (...) & Weinert, R. (1991). The HCRC map task corpus. *Language and speech*, 34(4), 351-366.

Bachman, L.F., Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5: 9/10, 341-345.

Bögels, S., Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52, 46-57.

Bosker, H.R., Pinget, A.F., Quené, H., Sanders, T. & De Jong, N.H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159-175.

Campbell, N. (2007). Approaches to conversational speech rhythm: speech activity in two-person telephone dialogues. *Proceedings of XVIth International Congress of the Phonetic Sciences*, Saarbrücken, Germany, 6-10 August 2007.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge, U.K: Press Syndicate of the University of Cambridge.

De Jong, N.H. (2016). Fluency in second language assessment. In Tsagari, D., Banerjee, J. (Eds.), *Handbook of second language assessment*. Mouton de Gruyter, 203-218.

De Jong, N.H., Groenhout, R., Schoonen, R. & Hulstijn, J.H. (2015). Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36(2), 223-243. DOI: 10.1017/ S0142716413000210.

De Jong, N.H., Steinel, M.P., Florijn, A.F., Schoonen, R. & Hulstijn, J.H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics,* 34(5), 893-916.

Derwing, T.M., Rossiter, M.J., Munro, M.J. & Thomson, R.I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54, 655-679.

Duncan, H., Segalowitz, N. & Phillips. N. 2014. Differences in L1 linguistic attention control between monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 19(1), 106-121. doi:10.1017/S136672891400025X.

Figueras, N., North, B., Takala, S., Van Avermaet, P., & Verhelst, N. (2009). *Relating language examinations to the common European framework of reference for languages: learning, teaching, assessment (CEFR): a manual.* Strasbourg, France: Council of Europe, Language policy division.

Fillmore, C.J. (1979). On fluency. In Kempler, D., Wang, W.S.Y. (Eds.), *Individual differences in language ability and language behavior*. New York: Academic Press, 85-102.

Grice, M., Savino, M. (2003). Map Tasks in Bari Italian: Asking Questions about Given, Accessible and New Information. Invited contribution to special issue *Intonation in Romance*. Catalan Journal of Linguistics 2.2, 153-180.

Hall, J.K. (1993). The Role of Oral Practices in the Accomplishment of Our Everyday Lives: The Sociocultural Dimension of Interaction with Implications for the Learning of Another Language. In *Applied Linguistics,* 14, 145-166.

Hall, J.K. (1995). (Re)creating Our Worlds with Words: A Sociohistorical Perspective of Face-to-face Interaction. In *Applied Linguistics,* 16, 206-232.

He, A.W., Young, R. (1998). Language proficiency interviews: A discourse approach. In Young, R. & He, A.W. (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency.* Amsterdam: John Benjamins, 1-24.

Jacoby, S., Ochs, E. (1995). Co-Construction: An Introduction. In *Research on Language and Social Interaction* 28, 171-183.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145- 164.

Kousidis, S., & Dorran, D. (2009). Monitoring convergence of temporal features in spontaneous dialogue speech. Dublin Institute of Technology: *Digital media Centre Conference Papers*. http://arrow.dit.ie/cgi/viewcontent.cgi?article=1003&context=dmccon.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. In *Behavior Research Methods*, 44, 325-343. URL http://www.lextale.com/.

Levinson, S. (2015). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences*. 20. 10.1016/j.tics.2015.10.010.

Logan, K.J. (2015). *Fluency disorders*. San Diego, CA: Plural.

McCarthy, M. (2009). Rethinking spoken fluency. *Estudios de lingüística inglesa aplicada,* 9, 11-29.

Magne, V., Suzuki, S., Suzukida, Y, Ilkan, M., Tran, M.N. & Saito, K. (2019) *Exploring the Dynamic Nature of Second Language Listeners' Perceived Fluency: A Mixed Methods Approach.* TESOL Quarterly. ISSN 0039-8322. https://doi.org/10.1002/tesq.528

Nava, E.A. (2010). *Connecting phrasal and rhythmic events: evidence from second language speech.* PhD dissertation. University of Southern California.

Préfontaine, Y., Kormos, J. & Johnson, D.E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language?. *Language Testing*, 33(1), 53-73.

R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes,* 14(4), 423-441. https://doi.org/10.1080/01638539109544795

Rossiter, M.J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, 65(3), 395-412. http://doi.org/10.3138/cmlr.65.3.395

Sacks, H., Schegloff, E.A. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.

Saito, K., Ilkan, M., Magne, V., Tran, M.N. & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, 39(3), 593-617.

Sbranna, S., Cangemi, F., Grice, M. (2019). Towards quantifying conversational competence in second language learning. Poster presented at *Phonetik und Phonologie Tagung*, Düsseldorf, Germany, 25-29 September 2019.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54, 79-95.

Segalowitz, N., Freed, B.F. (2004). Context, contact and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173-99.

Starkweather, C.W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice- Hall.

Suzuki, S. & Kormos, J. (2019). Linguistic Dimensions of Comprehensibility and Perceived Fluency: An Investigation of Complexity, Accuracy, and Fluency in Second Language Argumentative Speech. *Studies in Second Language Acquisition*. 10.1017/S0272263119000421.

Tavakoli, P. & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In Ellis, R. (ed.), *Planning and task performance in a second language*. Amsterdam: John Benjamins, 239-273.

Trouvain, J. & Truong, K.P. (2013). Exploring sequences of speech and laughter activity using visualisations of conversations. *Proceedings of Workshop on Affective Social Speech Signals* (WASSS 2013), Grenoble, 22-23 August, 2013.

Zmarich, C. (2017). Stuttering and phonetic theory: An introduction. In *Social and biological factors in speech variation,* studi AISV – Collana peer reviewed, 3, 357-372.