Multi-party Conversational Corpus of L1 and L2 for Speech Alignment Research (Teams-SK): Methodological Approach

Štefan Beňuš*†, Viktor Gatial*, Erik György*, Mária Hricková*, Martin Kažimír*, Zuzana Kozáčiková*, Lucia Mareková*, Robert Sabo†, Marian Trnka†, Erik Vráb*

*Constantine the Philosopher University in Nitra, Slovakia
†Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
{sbenus, vgatial, egyorgy, mhrickova, mkazimir, zkozacikova, lmarekova, erik.vrab}@ukf.sk
{Marian.Trnka, Robert.Sabo}@savba.sk

Abstract

The tendency for speakers to align or accommodate their verbal and non-verbal behaviour to their interlocutors is a fundamental mechanism in spoken interaction, strongly associated with successful communication and social bonding. Despite its ubiquity and documentation across various modalities and linguistic levels (e.g., lexical, prosodic), a lack of comparable, multi-layered linguistic resources and methodological agreement prevents a deeper understanding of its cognitive mechanisms. Multidimensional view of speech alignment might enhance its application in areas like language training or human-machine interaction. This paper addresses these gaps by presenting the development of a multilingual corpus of L1 Slovak and L2 English speech, extending a comparable corpus in L1 English. The corpus utilizes a modified cooperative board game, Forbidden Island, to elicit semi-spontaneous, multi-party conversation and introduces a complementary pair game to specifically target and prime syntactic alignment. The resource includes psychological metadata (e.g., personality, anxiety, perceived dominance) and enables a reproducible methodology for investigating the relationship between entrainment patterns and individual characteristics. By providing a non-Germanic language perspective and a direct L1–L2 comparison framework at prosodic, lexical, pragmatic and syntactic levels, this corpus offers a rich resource for advancing the theoretical understanding, replication, and practical application of speech alignment.

Keywords: Speech alignment, Multi-lingual corpus, L1 and L2 comparison

1. Introduction

The tendency of speakers to change their verbal and non-verbal behaviour and become more similar to their interlocutors has been extensively researched in various disciplines and is variably referred to as alignment, accommodation, or entrainment among others. Many studies documented the ubiquity of alignment in various forms of human spoken interactions, modalities, linguistic and paralinguistic aspects of speech (e.g. Rasenberg et al., 2020 for a review), and – particularly relevant for our paper – also in both L1 and L2 acquisition processes (e.g. Nielsenova and Swerts, 2012; Kim and Michel, 2023) or collaborative learning (Norman et al., 2022). Additionally, alignment has been associated with successful communication, positive social bonding (e.g. Beňuš, 2014), visual attractiveness (Michalsky, 2017), and many other aspects.

Multiple theoretical accounts and models have been proposed. According to Communication Accommodation Theory (CAT, Giles et al., 1991), people align to minimize the social distance between themselves and their interlocutor but may also dis-align to emphasize the social distance by acting less similarly. If spoken interactions, as a basic form of interpersonal contact, are seen as joint actions (Clark, 1996), speech alignment is a fundamental mechanism for "getting things done" in our lives, navigating social environments and for regulating our emotions. In the Interactive Alignment Model of Pickering and Garrod (2004), a production of a particular (linguistic) representation (e.g., word choice, syntactic structure, or a prosodic n intonational realization) automatically activates this representation in the mind of the conversation partner, which in turn increases the likelihood that the same representation will be reused by this partner.

Despite the ubiquity of alignment, theoretical research into the underlying roots, and its documented benefits for interpersonal communication, our understanding of its cognitive mechanisms, and the relationship to other aspects of human behaviour remains limited. These gaps prevent a wider application of alignment; for example in collaborative language/speech learning or human-machine conversational interactions.

Particularly, studies show alignment as a very complex phenomenon where linguistic (e.g. lexical, syntactic) and paralinguistic levels (e.g. speech rate, intensity, etc.) interact in non-trivial ways (Patel et al., 2022; Reichel et al., 2018; Wiese and Levitan, 2018). Moreover, the dynamic development of speech alignment, the relevance of dis-alignment, and how all this interacts with the interpersonal relationship among interlocutors are not well understood (e.g. de Looze et al., 2014). Furthermore, some studies also mention the lack of agreement on the methodological aspects of

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 6536–6546
11–16 May 2026.

alignment assessment. More specifically, it is not clear how to a) quantitatively measure various aspects of alignment and what the relationship among these aspects is (e.g. Kruyt et al., 2023), and b) how to best capture the discrepancies between alignment measured, or automatically extracted from spoken corpora using automatic processing tools on the one hand, and alignment qualitatively identified through close inspection and linguistic analyses of the collected data.

These gaps are partly related to the absence and fragmentation of easily available linguistic resources suitable for exploring alignment in both L1 and L2 in comparable settings, its links to the personality and emotional status of interlocutors, or potential identification of cross-linguistically valid vs. language/culture specific aspects of alignment. Moreover, this fragmentation leads to difficulties in relating the findings from different studies to one another and ultimately to the replicability issues in this line of research.

To address some of these issues, we extend an existing language resource in L1 English (Litman et al., 2016, see also below) by developing a comparable corpus of similar interactions in L1 Slovak—a West Slavic language—and in L2 English produced by the same Slovak speakers. In the following sections, we motivate and outline the methodological adjustments and additions as well as their rationale to facilitate the transparency and replicability of alignment research.

2. Available resources

Entrainment research relies on several established corpora, each emphasizing different interactional aspects. For example, The Brooklyn Multi-Interaction Corpus (Weise et. al., 2022) offers dyadic tasks supplemented with psychological and emotional metadata, valuable for linking alignment to individual differences. Additionally, The Teams Corpus (Litman et. al., 2016) captures four-party interaction through collaborative gameplay (*Forbidden Island*), making it one of the few resources to extend entrainment studies beyond dyads. It includes extensive team-related metadata, such as self-reports on personality, cognitive styles, and collective orientation. The task design also manipulated information gaps: half the teams received teamwork training, and novice players only had to collect three treasures instead of four. However, both corpora are restricted to L1 English only, with no multilingual and weak syntactic dimension, while the usefulness of the Teams Corpus for systematic L1–L2 comparisons is also limited.

Task-based corpora like the Columbia Games Corpus (Hirschberg et al., 2021), or Bielefeld Jigsaw Map Game Corpus (Lücking et al., 2011) are suitable for studying acoustic-prosodic alignment and also allow exploring lexical alignment or investigating turn-taking dynamics and referential communication. These tasks, however, often assign asymmetrical roles (instructor vs. follower), which, while suitable for isolating syntactic alignment, skews speech contribution and restricts opportunities for free alignment processes in natural dialogue.

Similarly, the Alternating Reading Task (ART) Corpus (Yuan et al., 2024) elicits alignment by having speakers take turns reading scripted material, and has been applied to cross-linguistic settings. The controlled design allows researchers to study phonetic convergence and the role of imitation in alignment, at the cost of spontaneous turn-taking, offering limited insight into pragmatic moves or syntactic entrainment in natural conversation.

The Spot the Difference (Lopes et al., 2016), Dialogues in Games (Boritchev and Amblard, 2022), and STAC (Asher et al., 2016) corpora highlight cooperation and negotiation but remain limited either by modality (STAC utilized written dialogue), monolingualism, or narrow focus on specific alignment mechanisms.

Multimodal datasets, such as RoomReader (Reverdy et al., 2022), D64 (Oertel et. al., 2013), and the Spot the Difference Corpus (Lopes et al., 2016) underscore the value of rich audiovisual data (e.g., D64 was recorded using several microphones and cameras tracking head, torso, and arm motion, the Spot the Difference Corpus also includes eye-tracking recordings, with rich annotations of verbal and non-verbal behaviour) and even relaxed conversational settings in entrainment research. Yet they do not easily lend themselves to extension in other languages.

One of the major gaps we thus identified is that most of the research on alignment, and also available speech corpora, has been conducted on native speakers of Germanic languages with some notable exceptions in Italian or Argentinian Spanish. Only a handful of studies compare alignment strategies of speakers among different L1s using comparable methodology (e.g. Levitan et al. 2015), and they find both similarities and differences. While understanding alignment in native language (L1) is important for theoretical modelling of the cognitive system of spoken interactions and potential future applied use, for example in human-machine spoken interactions (e.g. Beňuš, 2014; Dipold, 2022), understanding alignment in non-native (L2) spoken interactions and its links to L1 speech alignment has potential for better capturing the cognitive nature of this process. There is a recent study on speech alignment in L2 using speech processing methods (DeJong et al., 2022), but the scenario focuses on imitation rather than spontaneous production and a direct comparison with L1 is still missing. Moreover, as Pardo et al. (2018) suggest, alignment effects arising from less spontaneous and less interactive speech productions like

shadowing or imitation might not be directly comparable with speakers' alignment behaviour in more spontaneous and interactive settings.

Hence, what is missing are corpora that are neither too unconstrained to study systematic entrainment nor too restricted to capture the pragmatic and syntactic richness of dialogue. An audiovisual L1 Slovak L2 English corpus, combining spontaneous conversation with structured tasks, involving both multi-party interaction and dyadic tasks of the same speakers, and coupled with rich metadata that enables a better understanding of the role of team dynamics (e.g. dominance or various psychological aspects) in entrainment would fill several gaps outlined here. It would also enable direct L1-L2 comparisons at prosodic, lexical, pragmatic, and syntactic levels, and provide an additional non-Germanic perspective on entrainment. Finally, we believe that this resource will facilitate research into collaborative learning, applicable in the future to the L2 acquisition process within educational environments. Alignment might thus be useful for both screening and evaluating group interactions among learners, as well as serving as a potential intervention approach to make the collaborative learning process more effective.

3. Approach

3.1 Board game selection

Given the aim of exploring entrainment characteristics applicable to multilateral educational contexts, a choice to use a cooperative board game to provide a semi-structured experimental environment was made, partially on the basis of results obtained by Litman et al. (2016) and Paletz et al. (2023). *Forbidden Island* (FI) meets the base criterion of using a cooperative rule set. Nevertheless, preliminary tests were carried out with a number of other board games as well. These included non-cooperative games in which communication is necessary (e.g., *Catan*, *Munchkin*), as well as games with significantly shorter play time (e.g., *Codenames*, *Lascaux*). All tested games had to meet the criterion of having low rules complexity (evaluated on the basis of ratings provided in online fora and researchers' personal experience), accounting for test subjects with little to no experience with board games.

It was concluded that FI provides an optimal cooperative task environment at appropriate rules complexity. However, some adjustments to the game layout were necessary to reduce the play time – in contrast to Litman et al. (ibid.), the experiment includes more steps (notably an added pair game, see Section 3.3), and as such, the relatively long play time of FI under normal circumstances with novice players would have extended the duration of an experiment session beyond feasible time. A partial motivation for using FI is also the ability to build a future comprehensive resource. This resource might contain comparable L1 English speech (Litmann et al., 2016), newly added L1 non-Germanic speech, and L2 English, providing rich opportunities to explore alignment across multiple dimensions.

3.2 Game modifications

In order to facilitate feasible experiment session duration, increase the opportunities for entrainment, and increase the odds of subjects' success in the game, several adjustments to the game rules were made.

First, knowing that a research session will comprise pre- and post-experiment questionnaires, practice rounds, debrief discussions and an additional pair game, which also has to be played multiple times, it was decided that the play time of FI has to be reduced. This was achieved by minimizing the average number of actions a player has to take by the time the game ends by reducing the number of tiles used (while maintaining the movement-optimized square layout of the board), and decreasing the number of treasures the players have to capture as the game's objective to three.

Then, the difficulty and complexity of the game also had to be minimized so as not to lower the odds of success in the game and thus maintain conditions suitable for entrainment. The first adjustment was the full elimination of the fourth type of treasure card from the corresponding deck to reduce the number of choices the subjects have to make. To compensate, the number of remaining cards in the deck was increased to maintain the ratio of beneficial cards to the "waters rise!" event card, as the shift caused by removing the treasure card would greatly increase the game's difficulty. Second, the setup game rule was adjusted to reflect the reduced board size by having only four tiles flooded at the start of the game instead of the original six.

In order to maintain control over experimental conditions, specific pre-determined game setups (tile placement, deck orders, player role distribution) have been designed, so that each subject group experiences the game's input in the same manner. Subjects first play one trial round (intended to help verify their comprehension of the rules and serving as an unstructured teamwork practice), followed by two full games, played alternatively in Slovak and English in order to obtain data for comparative analysis of entrainment in native and foreign languages. The setup of the trial round is designed to increase the tension – and the perceived need for cooperation to achieve the defined goal – by featuring two "waters rise!" events in a short sequence; while the setups of the full games are structured to provide a lower-difficulty environment and allow

for winning the game in as little as ten rounds given relatively optimal play decisions. Rules on shuffling the individual discard piles were kept in order to maintain a sense of playing an actual board game rather than tackling a fully structured experimental task in order to help minimize Hawthorne effect (Marchionni et al., 2024).

Other adjustments of the game focused on increasing the affordances for entrainment. Game components were altered by removing lingual elements, notably location names from board tiles and corresponding flood cards, which was done to ease identification of the tiles the players are to interact with. This was coupled with an addition of a rule that the subject drawing flood cards is not allowed to flood (i.e., flip) the corresponding tiles themselves, but must describe the card's artwork to the other subjects, who are to identify the corresponding tile. This change increases the amount of different entities referred to in the conversation and simultaneously reduces systemic restrictions on lexical choice for the subjects. Crucially, it might also increase opportunities for observing entrainment at the syntactic level, as subjects' descriptions of tile art are expected to have a more complex syntactic structure than the pre-defined phrases provided on the original game components. Game components containing rules summaries (i.e., cards that trigger game events or allow players to perform special actions when drawn) were kept in their form with lingual elements, to account for players not being fully familiar with the rules at the start of the session.

To allow for observation of entrainment between persons in power-unequal social situations (as in educational environments between a teacher and a pupil), the role of the *Navigator* is always assigned to the research subject expected to hold a dominant position in the game sessions. This choice is based primarily on a pre-experiment ICL instrument – Interpersonal Diagnosis Questionnaire (Leary et al., 1976). (see section 3.4). However, it is expected that those subjects who have more experience with board games and/or show greater foreign language skills will be more at ease in the experiment and, at the same time, less experienced subjects will seek out guidance of the more experienced ones in order to meet the goal of winning the game. Therefore, subjects' observed degree of English proficiency and level of experience with board games are taken also into account when determining the expected dominant player. The role of *Navigator* was chosen for the dominant position due to its specific ability to move other player's pawns (i.e., the ability to impose their decisions on other players). To reinforce the perception of the *Navigator's* dominant position at the table, a written summary of rules is provided only to this player. This manipulation increases the expected amount of communication between the Navigator and other subjects, as the rules are now exclusively available through the *Navigator* to resolve other players' uncertainties.

The described implemented changes are also in line with those reported by Litman et al. (2016).The only notable difference is that the two game setups (one for L1 Slovak and one for L2 English game) used in our experiment are not isomorphic. This decision was reached because subjects might easily notice this isomorphism of setups, especially when drawing the "waters rise!" event cards and might thus predict the order of the tiles to be flooded.

3.3 Pair game addition

3.3.1 Rationale

Syntactic alignment (SA) is included in our study as part of a broader investigation of alignment across multiple linguistic levels. The analysis of syntactic alignment is particularly compelling because speakers are often unaware of the syntactic structures they produce or of the fact that they are mirroring each other. Moreover, unlike lexical or semantic alignment, syntactic alignment is not essential for efficient communication, making it a particularly subtle and implicit form of coordination. Finally, examining syntactic alignment enables cross-linguistic comparisons and contributes to a more comprehensive understanding of the mechanisms underlying interpersonal communication.

However, while the FI game offers suitable cooperative setting for studying alignment in word choice, prosody, and partly also pragmatics, the affordance for syntactic alignment in fully uncontrolled spontaneous settings is decreased. This arises through various reasons, primarily the common incompleteness of syntactic structures or strong tendency toward simplified structures disambiguated through visual and contextual factors. Settings with a more controlled interaction in pairs and based on resolving information gaps have more potential for the occurrence of syntactic alignment. This is reflected in previous studies which investigated syntactic alignment in pair activities (Branigan et al., 2007; Costa, 2008; Ostrand and Chodroff, 2021) that tested how interlocutors aligned their syntax to recently-preceding sentence. Commonly, studies tested mainly the use of the prepositional–object and the double-object dative sentences in native speakers' dialogues during picture description tasks (e.g. Branigan et al., 2000; Arai et al. 2007).

To address these needs, we designed a syntactic task that follows the FI game framework but also takes into account previous experimental studies on syntactic priming. Specifically, our setup follows the picture-matching paradigm that has been shown as the most effective environment to test SA (Branigan et al., 2020; Gruberg et al., 2019). Additionally, it provides material that will be

more comparable to these previous studies. Furthermore, inclusion of such a game would provide a complementary material not only for syntax, but also for prosody or turn-taking given more control over the spoken interaction.

Our syntactic experiment has two primary goals. Firstly, we aim to analyse and compare the differences in alignment between a native language (Slovak) and a non-native language (English), and, in line with this, to test which syntactic structures in English and Slovak are more likely to be primed.

3.3.2 Procedure

The experiment is carried out in pairs, with participants separated by an artificial divider screen so they cannot see each other's cards but can hear and see each other well. Each participant is given a playing grid with 20 tiles divided into two Sets corresponding to the rows of the grid and representing the two syntactic structures primed in that given set (e.g. Set 1 in the English game focused on passive constructions, while Set 2 targeted adjective order). The grids are seen on the top and bottom of Figure 1.

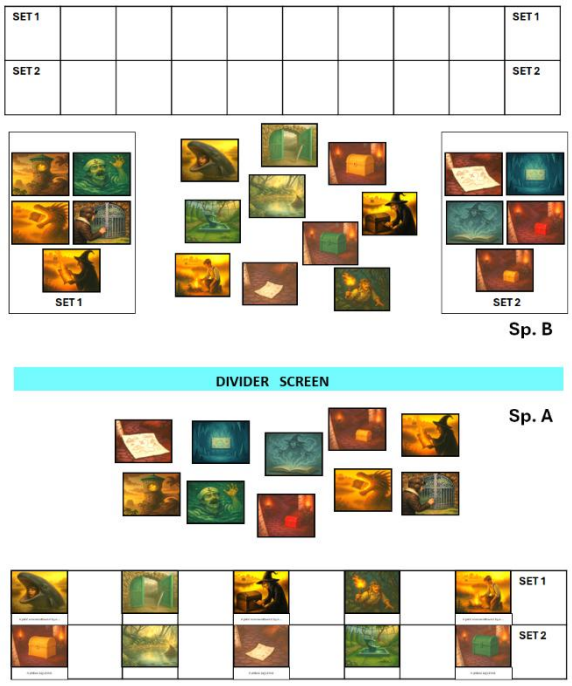

Figure 1: Boards in the pair game for Speaker A (Sp.A) bottom, and Speaker B (Sp. B) top

The procedure consists of two sequential phases: the Exposure Phase (where the subjects are exposed to syntactic production schemas) and the Test phase (where they are asked to produce their grammatical structure during a picture description task freely). In particular, during the Exposure Phase, Speaker A (bottom of Figure 1) describes pictures on their grid - half of them contain prescribed syntactic constructions in the form of text prompts on the bottom of the picture - the other half serve as distractors. Based on Speaker A's descriptions, Speaker B selects the correct pictures and places them to their own grid.

In total, participants play the game twice in their L1 (Slovak) and twice in their L2 (English). In each language, the roles switch for the second game - i.e. the participant who had prescribed syntactic structures (Speaker A) in the first game becomes Speaker B in the second game and no longer has the syntactic primes. The order of the languages is counterbalanced in the dataset. The four games (two Slovak and two English) are always preceded by a practice round in the language of the first upcoming game.

Eight sets of sentence primes are used in total (four in each language, two per game). These primes are based on preferred vs. dispreferred syntactic structures. Speaker A's cards always include the dispreferred syntactic structures. In Slovak, these include adjectival modifier order (e.g. *red small cabinet* vs. *small red cabinet*), participal constructions (e.g. *holding a compass the pilot is looking around the cave*), possessive constructions (e.g. *a sword of the messenger*), and fronted adverbial constructions (e.g. *in the cave there is a chest*). In English, the primed structures include also adjectival modifier order in the same way as in Slovak, di-transitive constructions (e.g. *A pilot is giving a sword to a girl*), locative alternations (e.g. *A girl is spraying the statue with perfume*), and passive structures (e.g. *A map was burned by a witch*).

In the experiment, each set represents a single syntactic structure, corresponds to one row on the boards shown in Figure 1, and contains two copies of ten different picture cards, one for each speaker. Both speakers describe five cards and place five cards. Of these 20 cards, ten per speaker, three contain a textual prompt eliciting the target syntactic construction, and the remaining ones have no textual prompt but can in principle be described with both preferred and dispreferred syntactic constructions of the particular set. The arrangement of the three primes and two non-prime targets for describing on Speaker A's board is fixed and identical across participants. When describing, Speaker B is free to select from the five non-prime targets in any order.

This procedure provides a novel and efficient scenario for testing alignment at syntactic level across two languages while keeping the task interactive and comparable to previous alignment studies.

3.4 Collection of psychological characteristics

Studies show that much of the variability in conversational data in general (Ivanov and Werner, 2010; Josserand et al., 2021) and also in entrainment patterns observed in the data (Lewandowski and Jilka, 2019) can be traced to

personality and other differences among individuals. Hence, we employed a set of questionnaires to capture some of these underlying factors. The surveys explored personality dimensions, anxiety, or perceptions of dominance of one's own and one's conversational partners.

Incorporating this information allows for a deeper interpretation of patterns in relationships, behaviours, and traits. By systematically integrating such background data, the resulting language resource will help establish a more robust basis for examining interactive behaviour better grasp the intricate processes involved in human communication.

Collection of psychological characteristics were carried out partly electronically (before the experiment) and partly on paper (during the experiment). We use the following psychodiagnostic instruments.

NEO-FFI is a standardized 60 item five-factor personality inventory that assesses an individual's level in five personality traits: neuroticism, extraversion, openness to experience, conscientiousness and agreeableness (Costa et al., 2007).

The ESQ-R Executive Skills Questionnaire-Revised (Strait et al., 2019) measures five areas of executive functions using 25 self-assessment items and consists of five subscales. The Planning subscale reflects logistical skills and the ability to complete scheduled tasks. The Time Management subscale refers to the ability to manage various aspects of time, including time estimation, time allocation, and the ability to work within time limits and constraints. The Organizing subscale refers to the ability to create and maintain systems for tracking information or materials. The Emotional Regulation subscale refers to the ability to manage emotions or control behaviour. Behavioural Regulation refers to the ability to exercise self-control and think before an individual acts or reacts, to consider the consequences of their actions.

The STAI X1/X2 (State-Trait Anxiety Inventory, Spielberger et al., 1999) is based on Spielberger's concept of distinguishing between anxiety as a state and anxiety as a personality trait and the possibility of measuring the differences between them, i.e. between a temporary, transient state and a relatively permanent predisposition. The overall concept of the questionnaire is based on the assumption that highly anxious people (with a high tendency to anxiety) will perceive situations or conditions that potentially imply the possibility of failure or self-harm with greater intensity than their less anxious counterparts.

To explore the relationship between entrainment and dominance, we decided to pre-select the most dominant member of the group and foster it by giving this person special power position of instructions bearer (see sections 3.2 and 4). We evaluate dominance of all subjects with eight items taken from the standardized ICL instrument – Interpersonal Diagnosis Questionnaire (Leary et al., 1976). The items correspond to the interpersonal type "autocratic personality", which is saturated with a combination of the "A" and "P" dimensions of the ICL questionnaire. The "A" dimension is characterized as dominance in social relationships and the "P" dimension as success recognized by other people.

After the board game played (in L1 and L2), the subjects fill out paper questionnaires that determine the level of perceived dominance and sympathy of the other teammates during the game play. The principle of the sociometric-rating questionnaire (SORAD) is used in the creation of the dominance and sympathy questionnaires (Gajdošová, 2000). This questionnaire was originally used to identify the informal ties of an individual in a social group formed by a school classroom.

3.5 Pre-experiment game instruction: video demonstration

An important addition to our research is the use of a Slovak-language video demonstration of the rules and gameplay, with rich visuals and narration reflecting our modified version of the game.

Commercial instructions were only available in Czech and did not account for our research-specific adjustments. Because Czech is closely related to Slovak, relying on those materials would have risked linguistic interference during the Slovak-language games. Translating the game components and instructions into Slovak ensured the integrity of the L1 data. A further motivation was the relative complexity of the game itself, which demands clear instruction.

Beyond addressing these practical needs, the video also served further methodological purposes: ensuring experimental control and enhancing data quality. By delivering identical Slovak instructions to all participants, it eliminates experimenter bias, ensures that everyone, even without prior exposure to the game, enters the task on the same footing, and guarantees consistency across sessions.

Figure 2: Screenshot from the video manual

To make the video manual, we first scanned all the game pieces, such as cards and tiles, and turned them into separate images. We also added additional visuals from royalty-free, copyright-free image sources, like the table with four players shown in Figure 2. Then we built each scene in PowerPoint, using our written manual as a guide for both the visuals and the voice-over script. We used PowerPoint's animation tools to make elements appear or fade in at the right time, highlight important parts with pulse and grow effects, and move cards along paths to show actions like drawing or discarding. Throughout the process, we aimed for simplicity in the visuals to avoid overloading viewers with too much sensory information. After finishing the animations, we recorded each scene as a short video and then combined them with the voice-over in video editing software to create the final video manual.

The multimodal format also reduced cognitive load, while the division of the video into shorter, self-contained sections with summaries provided an additional layer of comprehension support. This was implemented in BookWidgets (an online interactive learning platform), which allowed us to insert pauses after each video segment with a short written summary, as shown in Figure 3. Participants could also revisit earlier segments and review the summaries as needed. Providing participants with a prior understanding of the game mechanics directed the recorded interactions toward strategic collaboration and away from extended rule negotiation. Post-experiment interviews show positive responses of the subjects to the demo video verifying our intentions and reaching our goals.

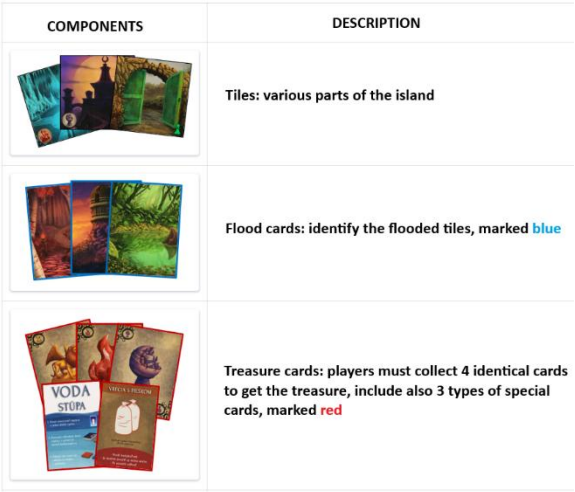

Figure 3: Example of a short summary in the interactive video manual; translated from original Slovak

4. Data collection procedure

The experiments take place in a sound-treated room equipped with Shure SM-35 head-mounted condenser microphones connected to Zoom H6 digital recorders and a 360° Kandao Meeting Pro conference camera. Audio and visual recordings are synchronized through initial hand claps.

Each session includes four Slovak adult participants who reported English as their second language at a self-perceived proficiency of at least B1. The groups are assembled to ensure that participants are unacquainted, as prior acquaintance could act as a confounding factor in the analysis of alignment. Other factors such as L2 proficiency or biological sex are not intentionally varied but we strive to produce as balanced groups as possible. In the first step, we recorded adults 22 and older, and collect the same number of sessions also for the young adults and teenager group (15-21 years). In total, we aim for 32 four-person sessions recorded, which is comparable to the four-person session recorded in the L1 English Teams corpus.

Figure 4: Real Forbidden Island game during data collection

The procedure is conducted in five stages and administered by two experimenters. In *Stage 1*, participants are introduced to the laboratory environment, asked to read a description of the upcoming procedure, and complete both an informed consent form and the State-Trait Anxiety Inventory (see section 3.4).

Stage 2 is designed to capture individual speech samples representative of the participants' neutral speaking styles. Participants are invited one at a time to the main recording room, fitted with a microphone, and calibrated for sound levels. Each participant then provides a detailed description of a picture for at least a minute. Because participants enter the room individually, they cannot hear the preceding descriptions, ensuring that their speech is not influenced by others.

In *Stage 3*, all four participants play a modified version of the collaborative board game *Forbidden Island*. The game is played in two full rounds, one in Slovak and one in English. The order of languages (SK-EN vs. EN-SK) is systematically counterbalanced across sessions to ensure equal representation in the corpus and

to maintain the comparability of future L1 and L2 alignment measures, given that the second game is often perceived by participants as easier than the first.

The full games are preceded by a practice round conducted in the same language as the first full game. The practice round familiarizes participants with the rules available in written form to all participants, which then can be discussed with the experimenter. The subsequent games are played without the written instruction and experimenter intervention. Gameplay is recorded continuously without breaks between the rounds so any exchange of game strategies or other relevant interaction between participants is captured. After each full game, the SORAD sociometric rating questionnaire (see section 3.4) is administered. Following the second game, a debriefing session takes place, beginning with pre-prepared open-ended questions and continuing as a free discussion.

Stage 4 consists of four rounds of the pair game (section 3.3), two in each language with the order counterbalanced to match the order in the *Forbidden Island* game. The full rounds are preceded by a practice round in the language of the first full round. Two participants are taken to a separate sound-treated room while the remaining two stay in the main laboratory room. The games are then played in parallel and take place as described in section 3.3. No breaks or questionnaires are administered during this stage. After the practice and four full rounds, a debriefing session takes place following the same procedure as in *Stage 3.*

Finally, *Stage 5* comprises the post-recording procedure. The participants are informed about the nature of the experiment and compensated for their time with vouchers.

5. Data processing

We begin by manually segmenting the continuous multichannel recording of each session into individual sections. The first section always consists of baseline recordings in which participants are asked to describe a card not used in the games (see *Stage 2* above). Each channel in this section therefore contains clean annotations indicating when a particular speaker is speaking. Based on these annotations, a speaker enrolment (i.e., a voice sample) is created for each participant.

Subsequently, automatic speaker diarization is performed separately for each channel. We use the Diarizen toolkit (Han, 2025), with the model BUT-FIT/diarizen-wavlm-large-s80-md, which yielded the best results in our experiments. During this process, the detected speech segments are verified against the enrolled voice sample of the speaker expected to appear in the given channel. This procedure allows us to identify, and disregard, instances of speech from other speakers present in the processed audio channel.

The final output is a Praat TextGrid file with separate tiers for each speaker derived from four independent diarization processes corresponding to the individual channels. Based on this diarization, automatic speech recognition is then performed for each channel. We use the OpenAI Whisper library with the CrisperWhisper model (CrisperWhisper) for both Slovak and English sessions, and the recognized text is written into the corresponding TextGrid tiers.

The quality of the initial transcriptions was not assessed quantitatively; however, repeated evaluations of several data subsets indicated that ASR-generated transcripts provide a more efficient starting point for annotators than transcribing from scratch. At the same time, the quality of the ASR output was not sufficient and required human correction. The final step thus involves manual correction of transcription errors by an expert. The final transcripts include continuers like *uh-huh* or *mhm*, conversational fillers like *uh*, *um*, and *mm*, laughs, incomplete words (marked with a hyphen), and unrecognized words or phrases (marked as “?”). The quality of the post-corrected aligned transcriptions is ensured through repeated checks of random sections of the recordings to verify the placement of boundaries defining interpausal units (IPUs) and the accuracy of the transcription.

The type and extent of annotation required to address specific research questions related to alignment will necessarily depend on the analytical focus of individual studies (e.g., lexical, prosodic, syntactic, or interactional phenomena). For this reason, the corpus will be made available with time-aligned transcriptions at the level of IPUs, while further linguistic annotation will be left to researchers according to their particular methodological needs. This design choice ensures flexibility and broad usability across different research agendas.

6. Corpus structure and availability

The speech dataset is organized using the schematic directory hierarchy below. The .wav files contain multichannel audio recordings and .TextGrid files the transcriptions in tiers corresponding to the speaker channels. The file names include the game (FI for Forbidden Island and PG for Pair game), speakers (s1234 for FI games and s12, s34 for PG games), and language (sk for Slovak and en for English). Metadata for each session include information about the age, sex, L2 proficiency, collected personality and psychological information (section 3.4), etc.

A sample of the future corpus is available at https://speech.savba.sk/TEAMS_SK/. After the completion of recording and processing, the

corpus, together with the video demonstration and selected metadata will be available through ELRA (European Language Resources Association).

```
dataset_root/
├── session_01/
│   ├── baseline_s1234_sk.wav
│   ├── game_FI_s1234_sk.wav
│   ├── game_FI_s1234_sk.TextGrid
│   ├── ...
│   ├── game_PG_s12_en.wav
│   ├── game_PG_s12_en.TextGrid
│   ├── ...
├── session_02/
│   ├── ...
└── metadata
```

7. Conclusion

This paper introduces the design of a new bilingual, multi-party conversational corpus developed to facilitate analyses of speech alignment in L1 and L2 contexts. This corpus extends existing L1 English datasets (e.g., Litman et al., 2016) with comparable Slovak L1 and L2 English data and incorporates a novel syntactic priming game built on a cooperative board task. This approach provides an opportunity to examine speech alignment across prosodic, lexical, pragmatic, and syntactic dimensions. For example, researchers may examine how native language (L1 vs. L2) influences the nature and dynamics of lexical alignment when establishing how to name game cards. Furthermore, the inclusion of measures of dominance, personality, or anxiety enables exploration of how interpersonal dynamics shape speech alignment. For example, researchers may test hypotheses concerning the tendency for speech alignment to be positively correlated with anxiety and negatively correlated with dominance.

The design of this corpus bridges the methodological and resource gaps identified in current research and provides a foundation for future cross-linguistic and multimodal studies that might further illuminate the cognitive and social mechanisms underlying human communication. Beyond these research applications, the corpus lays the groundwork for practical use in language education, group communication, and the design of adaptive conversational technologies (e.g. human-machine communication). In particular, understanding how dominance shapes alignment can help teachers use their natural communicative authority to foster positive alignment and support less proficient speakers. In human-machine interaction, the observed alignment patterns can inform the design of adaptive personalized systems that would be capable of more natural and perceptually pleasant communication.

8. Ethical considerations

So far, 80 participants (20 sessions) were recruited, signed a consent to record their audio and video and analyse the interactions by the experimenter's institutions for academic purposes, and distribute the recorded anonymized speech to other researchers. We plan to collect data from 128 participants.

9. Limitations

Though carefully planned, the current design of data collection has several limitations. While efforts were made to avoid predictability in the game tasks, the resulting lack of uniformity may reduce comparability of L1 (Slovak) and L2 (English) samples. Similarly, the inclusion of participants from different age groups may require distinct approaches in future recordings (and potentially in data analyses). Such stratification within the final corpus could also limit the statistical power of certain analyses. This limitation may be further influenced by a relatively high number of psychological factors within a relatively small sample, which might result in a considerable variety of observed psychological patterns. Finally, the induced dominance manipulation, although methodologically justified, represents only one type of social hierarchy and may not fully capture the complexity of natural group dynamics.

10. Acknowledgments

This research is supported by the Slovak Research and Development Agency, Slovak Republic grant APVV 23-0223 and the Slovak Granting Agency grant VEGA2/0092/25.

11. Bibliographical References

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Memory & Cognition,* 35(7), 1789–1802.

Beňuš, Š. (2014). Social Aspects of Entrainment in Spoken Interaction. *Cognitive Computation* 6, 802–813.

Bortfeld, H., and Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes, 23*(2), 119–147.

Branigan, H.P., Pickering, M.J.& Cleland, A.A. (2000). Syntactic coordination in dialogue, *Cognition* 75, 13-25.

Branigan, H.P. et al. (2007). Syntactic alignment and participant role in dialogue, *Cognition* 104, 163-197.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Costa, P. T., McCrae, R. R., Ruisel I., & Halama, P. (2007). NEO päťfaktorový osobnostný inventár, 1. slovenské vydanie, [NEO 5-factor peronality inventory, 1st Slovak edition] Praha: Testcentrum-Hogrefe.

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes,* 23(4), 528-556.

CrisperWhisper https://github.com/nyrahealth/CrisperWhisper

De Jong, D., Pastore, A., Nguyen, N., D'Ausilio, A. (2022) Speech imitation skills predict automatic phonetic convergence: a GMM-UBM study on L2. Interspeech 2022, 769-773.

De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication* 58, 11-34.

Dipold, D. (2022). Making the case for audience design in conversational AI: Rapport expectations and language ideologies in a task-oriented chatbot. https://arxiv.org/abs/2206.10694.

Gajdošová, E. (2000). *Poznávanie sociálnych vzťahov v triede. Práca so sociometricko-ratingovým dotazníkom* [Learning about social relationships in the classroom. Work with a sociometric-rating questionnaire]. Bratislava: Metodické centrum.

Garrod, S., & Pickering, M. J. (2007). Alignment in dialogue. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 443–452). Oxford University Press.

Giles H., Coupland N., & Coupland J. (1991). Accommodation theory: Communication, context, and consequence. In: Giles H. Coupland N. Coupland J, (eds). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press, pp. 1–68.

Gruberg, N., Ostrand, R., Momma, S., Ferreira, V.S. (2019). Syntactic entrainment: The repetition of syntactic structures in event descriptions, *Journal of Memory and Language*, 107, 216-232.

Han, J., Pálka, P., Delcroix, M., Landini, F., Rohdin, J., Cernocký, J., & Burget, L. (2025). Efficient and Generalizable Speaker Diarization via Structured Pruning of Self-Supervised Models. *arXiv preprint arXiv:2506.18623*.

Ivanov, M., & Werner, P. D. (2010). Behavioral communication: Individual differences in communication style. *Personality and Individual Differences*, 49 (1), 19–23.

Josserand, M., Allassonnière-Tang, M., Pellegrino, F., & Dediu, D. (2021). Interindividual Variation Refuses to Go Away: A Bayesian Computer Model of Language Change in Communicative Networks. *Frontiers in Psychology*, 12.

Kim, Y., & Michel, M. (2023). Linguistic alignment in second language acquisition: A methodological review. System, 115, Article 103007.

Kruyt, J., de Jong, D., D'Ausilio, A., & Beňuš, Š. (2023). Measuring Prosodic Entrainment in Conversation: A Review and Comparison of Different Methods. *Journal of speech, language, and hearing research : JSLHR*, 66(11), 4280–4314.

Leary, T., LaForge, R. L., Suczek, R. F. Kožený, J., & Ganický, P. (1976). *ICL – Dotazník interpersonální diagnózy. Příručka pro administraci, interpretaci a vyhodnocování testu* [Interpersonal check list - Interpersonal Diagnosis Questionnaire. Manual for administration, interpretation and evaluation of the test]. Bratislava: Psychodiagnostické a didaktické testy.

Levitan, R., Beňuš, Š., Gravano, A., & Hirschberg, J. (2015). Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. *Proceedings of SIG-DIAL*, pp. 325-334.

Lewandowski, N., & Jilka, M. (2019). Phonetic Convergence, Language Talent, Personality and Attention. Frontiers in Communication, 4.

Litman, D., Paletz, S.,Rahimi, Z., Allegretti, Z., & Rice, C. (2016). The Teams Corpus and Entrainment in Multi-Party Spoken Dialogues. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1421–1431, Austin, Texas, November 1-5, Association for Computational Linguistics.

Marchionni C., Zahle J., & Godman M. (2024). Reactivity in the human sciences. *European Journal for Philosophy of Science, 14*(1), art. no. 8.

Michalsky, J., Schoormann, H. (2017) Pitch Convergence as an Effect of Perceived Attractiveness and Likability. *Proc. Interspeech 2017*, 2253-2256.

Nilsenova, M. & Swerts, M. (2012). Prosodic Adaptation in Language Learning. In J. Romero-Trillo (ed.), Pragmatics and Prosody in English Language Teaching. Dordrecht: Springer.

Norman, U., Dinkar, T., Bruno, B., & Clavel, C. (2022). Studying Alignment in a Collaborative Learning Activity via Automatic Methods: The Link Between What We Say and Do. *Dialogue & Discourse*, 13(2), 1-48.

Ostrand, R., & Ferreira, V. S. (2019). Repeat after us: Syntactic alignment is not partner-specific. *Cognition*, 192, 104012.

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features, but not others within a dialogue. *Journal of Phonetics*, 88, 1–17.

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., et al. (2018). A comparison of phonetic convergence in

conversational interaction and speech shadowing. *Journal of Phonetics* 69, 1–11.
Paletz, S. B. F., Litman, D., Karuzis, V., Jones, K. M., & Rahimi, Z. (2023). Speaking Similarly: Team Personality Composition and Acoustic-Prosodic Entrainment. *Small Group Research*, 54(6), 860-898.
Patel, S. P., Cole, J., Lau, J. C. Y. et al. (2022). Verbal entrainment in autism spectrum disorder and first-degree relatives. *Scientific Reports* 12, 11496. doi.org/10.1038/s41598-022-12945-4.
Pickering, M.J.& Garrod.S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2), 169-190.
Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. Cognitive Science, 44(11), e12911.
Reichel, U., Beňuš, Š., & Mády, K. (2018). Entrainment profiles: Comparison by gender, role, and feature set. Speech Communication 100, 46-57.
Spielberger, D., Gorsuch, R., Lushene, B., & Ruisel, I. (1999). Dotazník na meranie úzkosti a úzkostlivosti STAI [Questionnaire for evaluating state-trait anxiety]. Bratislava: Psychodiagnostika a. s.
Strait, J. E., Dawson, P., Walther, C. A. P., Strait, G. G., Barton, A. K., & McClain, M. B. (2019). Rrefinement and psychometric evaluation of the executive skills questionnaire-revised. Retrieved from: https://www.scribd.com/document/447289137/executive-skills-questionnaire-revised-esq-r
Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. In *NAACL: HLT Proceedings, Volume 2 (Short Papers)*, pp. 297-302.

12. Language Resource References

Asher, N., Hunter, J., Morey, M., Benamara, F., & Afantenos, S. 2016. Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus. In *The Tenth International Conference on Language Resources and Evaluation* (LREC 2016), pp. 2721–2727.

Boritchev, M., & Amblard, M. 2022. A Multi-Party Dialogue Resource in French. arXiv:2207.12162 [cs.AI]. Available at: https://arxiv.org/pdf/2207.12162

Hirschberg, J., Gravano, A., Benus, S., Ward, G., & Sneed German, E. 2021. Columbia Games Corpus. Linguistic Data Consortium, distributed via LDC: LDC2021S02, 1.0, ISLRN 834-843-130-497-9.

Lopes, J., Hemmingsson, N., & Åstrand, O. 2018. The Spot the Difference Corpus: a multi-modal corpus of spontaneous task oriented spoken interactions. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018). European Language Resources Association (ELRA).

Lücking, A., Menke, P., Abramov, O., & Mehler, A. 2011. The Bielefeld Jigsaw Map Game (JMG) Corpus. In Proceedings of the Corpus Linguistics 2011 Conference, Birmingham, UK. Available at: https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-137.pdf

Oertel, C., Cummins, F., Edlund, J., Wagner, P., & Campbell, N. 2013. D64: A Corpus of Richly Recorded Conversational Interaction. Journal on Multimodal User Interfaces, 7, 19–28. https://doi.org/10.1007/s12193-012-0108-6

Reverdy, Justine, O'Connor Russell, Sam, Duquenne, Louise, Garaialde, Diego, Cowan, R., Benjamin and Harte, Naomi (2022). RoomReader: A Multimodal Corpus of Online Multiparty Conversational Interactions. In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, pp. 2517-2527.

Yuan, Z., de Jong, D., Beňuš, Š., Nguyen, N., Feng, R., Sabo, R., Fadiga, L., & D'Ausilio, A. 2024. ART: The Alternating Reading Task Corpus for Speech Entrainment and Imitation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING 2024), 1548–1562.