

Starting a Conversation with Strangers in Virtual Reykjavik: Explicit Announcement of Presence

Stefán Ólafsson^{1,3}, Branislav Bédi², Hafdis Erla Helgadóttir¹,
Birna Arnbjörnsdóttir², Hannes Högni Vilhjálmsson¹

¹CADIA, Reykjavik University

²University of Iceland

³Relational Agents Group, Northeastern University

stefanola13@ru.is, brb19@hi.is, hafdis13@ru.is, birnaarn@hi.is, hannes@ru.is

Abstract

Virtual Reykjavik is an Icelandic language and culture training application for foreigners learning Icelandic. In this video game-like environment, the user is asked to solve given tasks in the game and in order to complete them he/she must interact with the characters, e.g. by conversing with them on context-specific topics. To make this a reality, a model for how natural conversations start in a specific situations has been developed, based on data from that same situation in real life: a stranger asking another stranger for directions to a particular place in downtown Reykjavik. This involved defining a multimodal annotation scheme, outlining the communicative functions and behaviors associated with them. However, current annotation schemes lacked the appropriate function for this specific case, which lead us to finding and proposing an appropriate communicative function – the Explicit Announcement of Presence. A study was conducted to explore and better understand how conversation is initiated in first encounters between people who do not know each other. Human-to-human conversations were analyzed for the purpose of modelling a realistic conversation between human users and virtual agents. Results from the study have lead to the inclusion of the communicative function in the human-to-agent conversation system. By playing the game the learners will be exposed to situations that they may encounter in real life, and therefore the interaction is based on real life data, rather than textbook examples. We believe that this application will help bridge the gap from the class room to the real world, preparing learners to initiate conversations with real Icelandic speakers.

Index Terms: Explicit Announcement of Presence, communicative function, human-agent interaction, embodied conversational agent, multimodal communication, natural language, social behavior

1. Introduction

The Icelandic language and culture training application *Virtual Reykjavik* is an on-line computer game environment supporting game-based learning [1], task-based learning [2] and a communicative approach [3, 1] to teach Icelandic as a foreign language (adult learners living outside of Iceland) or second language (adult learners living in Iceland). Learners (users from now on) can gain particular linguistic and cultural skills by engaging in interactive exercises and are then able to use that knowledge in conversations with real people in the natural setting of the target language [4]. The exercises entail practicing saying words and



Figure 1: A screenshot of an ECA in Virtual Reykjavik. A yellow arrow appears overhead when the user has targeted the ECA and the mouse may be clicked to activate speech recognition. The green, yellow, and red lights in the upper right corner indicate the user's changing role in the interaction, i.e. listener or speaker.

phrases in simple conversations with *Embodied Conversational Agents* (ECAs), which are defined as a computer interface represented by a humanoid body that is specifically conversational, exhibiting and recognizing the behavior involved during human face-to-face conversation [5]. In these interactions, the users find themselves in various situations, such as encountering a stranger, starting a conversation, and asking him/her for directions (see Figure 2).

In Virtual Reykjavik, users interact with different ECAs (male or female) in the following ways: by approaching an agent until it acknowledges the user's presence, using the mouse to signal which direction or which ECA the user is looking at, clicking the mouse to trigger an action, such as speaking, and by talking into the microphone through which the ECA gets the speech input from the user (Figure 1). Similar to the Tactical Language and Culture Training System [4], Virtual Reykjavik also relies on natural spoken language when interacting with game characters, i.e. ECAs, via automatic speech recognition. For Icelandic, the current version of our system uses the



Figure 2: The first image (from the left) shows the ECA's reaction to the user performing behavior associated with the Explicit Announcement of Presence (EAP), the second shows the agent's reaction to a question, the third shows the agent answering the question, and the last image shows the reaction to the user saying thank you.

Google speech recognition service¹, because it is currently the only readily available software for the language [6]. In the first learner scenario or level of the game, users have to fulfill three tasks: 1) to get an ECA's attention and start a conversation, 2) to ask the ECA for directions, and 3) to say goodbye to the ECA (see Figure 2). In this article, we discuss the finding and implementation of an appropriate communicative function associated with getting a stranger's attention and the multimodal behavior associated with the acknowledgement of it.

2. Motivation

In situations where participants know each other, a greeting phase often fulfills the function of noticing and acknowledging one's presence and initiating a conversation [7]. During exploratory data collection for the Virtual Reykjavik project, we found that when strangers approach one another and start a conversation, something other than a greeting occurs. We were motivated to take a closer look at what behaviors native speakers of Icelandic exhibit and what communicative functions they carry out, in order to provide the users of Virtual Reykjavik with an accurate portrayal of conversations in the language.

We recorded and annotated naturally occurring human-to-human conversations in order to faithfully emulate conversational behavior in Virtual Reykjavik. During the annotation process, we felt that the communicative function being conveyed at the very beginning of each encounter was not to be found in current standard multimodal annotation schemes, such as the MUMIN coding scheme [8], SmartKom multimodal corpus [9] or the HuComTech multimodal corpus annotation scheme [10].

In the theoretical exploration of this topic, we came upon the *Explicit Announcement of Presence* (EAP) [7], which we used for building our hypothesis: when strangers meet during first encounters in situations when they ask for directions, the Explicit Announcement of Presence is the communicative function underlying the behavior at a start of a conversation.

3. Modelling Approach

In order to design a realistic conversational structure for our context-specific situation using the Icelandic language, we needed to address two general problems. First, we needed to define the appropriate communicative functions and behaviors that would best fit our context, i.e. how a stranger (non-native

speaker) approaches another stranger (native speaker) in downtown Reykjavik, and how the approached person acknowledges it. We then needed to implement the function involved in getting someone's attention and the natural multimodal behavior associated with acknowledging it in the virtual characters. A traditional greeting-phase, often used as initial learning scenarios in textbooks, would not apply in the situation we picked, because traditional greetings are primarily used amongst persons who know each other. On that account we needed to come up with something new.

We worked towards realizing a conversational structure that would maintain presence and authenticity, with the aim of giving the user a feeling of a natural conversation akin to what we observed in our field study. The approach we took was inspired by Clark's (1996) *conversation sections*, which are purpose-specific segments of a conversation that arise during the course of face-to-face interaction between humans. On a very high level, these sections include the *entry*, *body*, and the *exit* of the conversation [11]. However, thinking of the body of a conversation as one single purpose-specific section is rather vague. Thus, for our purposes, we tried to identify portions of the conversations in our data as being potential conversation sections, portions where the participants are bound to an identifiable purpose.

An example of such a segment in a real-life context is during the task of asking a question and receiving an answer. The initiating participant has a purpose, i.e. to gain some knowledge from the other party, and in the process he/she alters the intent of the other. In other words, the initiator influences other participant's intentions and together they become involved in this purpose-specific segment, or conversation section.

The emergence of a conversation section at any given time during a conversation is governed by multiple factors, such as the relations between the participants, their intention, and personality. Moreover, all of these factors affect what functions and behaviors are involved in the context of the particular conversation section at hand. For instance, in an informal setting where participants know each other, a greeting would sound and look different to one in a formal setting where participants do not know each other. We defined the EAP as the appropriate function for initiating conversation in the following setting: (1) participants who are strangers; (2) a non-native speaker approaches a native speaker in an informal setting, which is downtown Reykjavik, and asks for directions to a particular place.

¹ <http://goo.gl/eSRnbv>

Track Type	Function Category	Type
Interactional	Initiate	react, recognize, salute-distant, salute-close, initiate
	Turn-taking	take, give, keep, request, accept
	Speech-act	eap*, inform, ask, request
	Grounding	request-ack, ack, repair, cancel

Table 1: These are the in interactional function categories from the original FML proposal [12] for use in the Virtual Reykjavik system (alteration marked with *).

4. EAP Study

4.1. Method

In order to better understand the use of EAPs for initiating conversations, both in terms of frequency of use and how they are manifested in behavior, we conducted a small qualitative study. Natural language data from conversations of first encounters was collected in the form of video recordings. The focus was on approaching a stranger and starting a conversation. Two volunteer actors, both female non-native speakers of Icelandic, were hired to approach Icelanders and ask for directions to a particular place in downtown Reykjavik. The first human-to-human conversations we recorded were done by walking up to people and stating our purpose beforehand. This made it impossible to capture the initial moments of naturally occurring contact. We therefore changed our method to stating the purpose of our research to people after the conversation. The actors received only one instruction: to ask people for directions. Without further telling the actors what to do, they started naturally approaching people and announcing their presence. Consent from participants was recorded on camera at the end of each recording and participants could ask to withdraw from the study and their recording would be deleted on the spot.

The actors were asked to conduct themselves as normally as possible. The effect of them being non-native speakers is negligible in these circumstances, since in all cases they performed the appropriate utterances and had clear pronunciation. The selective sampling method [13] was applied here in order to address the right group of people and ensure the authenticity of the collected data. Only male and female native speakers of Icelandic aged between 18-70+ were considered. The study was anonymous and concession was received from all participants.

The video recordings were annotated using a multimodal annotation scheme for Virtual Reykjavik compiled from various other research (see Figure 3), both in terms of the communicative behavior present in the dialogue and the underlying intent or function of those behaviors [14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. This follows the distinction between function and behavior made in the SAIBA framework for multimodal generation of communicative behavior for ECAs, as manifested in the Behavior Markup Language (BML) [18] and Function Markup Language (FML) [12]. Our current work contributes to existing work on FML by introducing the EAP as a type of a communicative function. In our observations, a verbal behavior typically follows the EAP. We therefore categorize the EAP as a type of *speech-act*, which is a communicative function category that includes multiple types (*ask*, *inform*, etc.), adding to the current FML standard (see Table 1).

4.2. Results

We analyzed 44 videos that included first encounters between native speakers and non-native speakers of Icelandic asking for directions to a specific place in downtown Reykjavik. The fo-

Multimodal Data for Modeling Agent's Response to EAP	
SPEECH.....	--
HEAD.....	Central, directed at speaker / user
FOREHEAD.....	Slightly crumpled
EYE BROWS.....	Slightly raised, slightly drawn together
GAZE.....	Direct, open eyes
MOUTH.....	Slightly open
HANDS.....	Beside the body, no movement
BODY POSTURE.....	Directed at the speaker / user, aligned with the torso
DISTANCE.....	Close to the speaker / user
TORSO.....	Directed at the speaker / user, aligned with the whole body
LEGS / FEET.....	Finish movement (walking), legs slightly apart when standing still

Figure 3: An example description of behavior using an annotation scheme developed for Virtual Reykjavik.

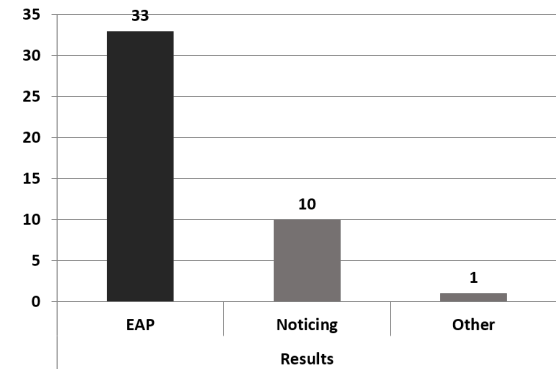


Figure 4: Most people starting an interaction with a stranger in the street (33 out of 44 videos) show behavior that carries out an EAP function.

cus was on the first part of the dialogue, i.e. approaching a person and initiating a conversation. The data shows that in 33 videos (75% cases) pedestrians passing through (non-natives) announce their presence verbally to other unknown pedestrians (natives), in 10 videos (23% cases) both notice each other before an announcing phase has a chance to happen, and in 1 video (2% cases) the phase was described as “other” because it could not be identified (see Figure 4).

Results show that in most of the cases approaching pedestrians announce their presence verbally in order to cause attention from the approached pedestrian to initiate a conversation about getting directions to a particular place. It became clear that a particular communicative function – the EAP – was primarily being conveyed verbally here by the non-natives when approaching the natives.

The most frequent EAPs in our data have the following form: 1) phrases: *fyrirgefðu* [pardon me], *afsakið* [excuse me], 2) greetings: *góðan daginn* [good day] with definite article / *góðan dag* [good day] without definite article / *hæ* [hi], or 3) directly asking the question: *Veistu hvar X er?* [Do you know where X is?]. In our study, explicit nonverbal EAPs were not

found, except when looks of approaching participants accidentally met. But this has been categorized in FML as *noticing*, because it involves a stranger (actor) gazing at another stranger (native speaker), who responds by gazing back and awaiting some kind of a response from the gazing actor. All of this takes place in a fraction of a second.

After the EAP is performed by the non-native stranger, the person being approached generally realizes that someone wants to speak to him/her and looks back at that person. Detailed analysis of the multimodal behavior, exhibited by the native speakers as a result of the EAP by the non-native speakers, was performed using representative subjects: one male and one female native speaker, both around 50 years of age. The data was annotated in Elan [24], and the Multimodal Annotation Scheme for Virtual Reykjavik was used as a reference. Results are listed in Table 2 and Table 3.

These results lead to the incorporation of the EAP communicative function into the Virtual Reykjavik ECAs, including the realization of plausible EAP related behavior.

5. Implementation

The EAP is just one of many communicative functions, among others such as *turn-taking* and *grounding*, that precede a set of one or more behaviors. The communicative plans of an ECA are manifested in communicative functions, i.e. pieces of intent that have a communicative purpose. Based on these functions, the system then plans out which behaviors carry them out.

While the functions themselves are unseen, the behavior is their visible result. If a person wants to approach another person who is a stranger, their brain plans for an EAP and when the time comes it tells the body to perform the behaviors associated with that function, e.g. to look at the other participant and say *afsakið* [excuse me].

Implementing the EAP function within the conversational system architecture of Virtual Reykjavik lets the users interact with the ECAs in a more realistic way, and the ECAs get to exhibit realistic behavior in response to it. This behavior also relates directly to one of the important tasks that users have to perform in order to fulfill the game objectives, i.e. engage with a stranger in the street and ask him/her for directions.

Our implementation involved the use of conversation sections. Within our system, these sections are called blocks and are the objects that contain methods for producing communicative functions that underlie the behavior in various situations [25]. Knowledge regarding which behaviors and functions are appropriate for each situation was gathered from the annotated video data (see Figure 3 and Tables 2 & 3).

As mentioned above, the *entry* is the first purpose-specific segment of the conversation and a block in our system that corresponds to that is the *Approach* block (see Figure 5). This block, based on the observed data, necessarily includes the EAP in order for the stranger to initiate a conversation with an ECA [25].

The block element structure allows the Virtual Reykjavik conversation system to procedurally select what comes next in the conversation. The blocks provide a context for the communicative functions at any given moment and align speech with other modalities, in our case the conversational behavior of the ECAs. We had to design a system architecture that allows the agent to make a decision as to what should happen next in the conversation, based on dialog history, personality, and what events have unfolded in the interaction with the user at any given time.

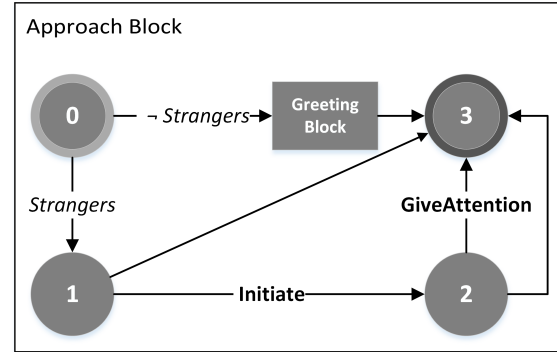


Figure 5: The ‘approach’ block’s state machine propels the conversation using methods (*Initiate* and *GiveAttention*, shown in bold) that generate discourse functions relative to the agents’ intent. The initial state checks for relations and moves to either a greeting phase or a ‘stranger specific’ initiation of conversation. States (1) and (2) allow for ‘inaction’, resulting in the approach coming to an abrupt end in the final state (3). [25]

This system allows the agents to either move to the next state within a particular block or, if the current block is finished, select which block of conversation they want to push next to the floor of interaction.

5.1. Initiating a Conversation

The following provides a more in-depth description of what transpires when the user approaches an agent in the first Virtual Reykjavik game scenario. When the user starts the program he/she embodies an avatar that is structurally very similar to the other ECAs in the scene, i.e. they both have perception systems that function in the same way and they perform behaviors in the same manner. The only difference is that the human user is in control of his/her avatar’s head movement, where he/she walks, and does certain actions with the keyboard and the mouse.

When the user (player) moves his/her avatar closer to the agent, their respective perception systems perceive the other and their reasoning faculties check their intentions in order to decide whether to act on them (see Figure 6). While the agent has no interest in initiating a conversation, the player’s intent for getting the information is made known by clicking the mouse when a yellow arrow appears above the agent’s head, as in Figure 1. This prompts the speech recognition software to allow the player to speak and when he/she is done speaking the speech recognition automatically stops listening. The input is stored for further analysis and may have an impact on which block will be selected next.

Following this action, the program instantiates a *discourse manager* and a *floor* of interaction is created with the agent and player as participants. The discourse manager asks the floor to execute the next action in the current block; however, in this case, it finds that no current block is available. Therefore, the first block is established by looking at both participants’ intentions and personality parameters and in this case an Approach block is selected.

The player and agent’s relationship is checked in the first state of the Approach, and here they are found to be strangers. The participants then progress to the next state where the player’s avatar creates a bundle of communicative functions

Nonverbal Reaction of B to A's EAP		Description
Head		central, directed at A
Face	Forehead	crumpled
	Eyebrows	slightly raised & slightly drawn together
	Eyes	open & directed at A
	Mouth	slightly open
Torso		slightly turned away from the A
Hands		beside the body, holding hands, no movement
Body posture		aligned with the torso = slightly turned away from the A due to A's interfering form the side
Position		close to the A

Table 2: Sample nonverbal reaction of female native speaker of Icelandic (B) to the EAP of approaching female non-native speaker (A)

Nonverbal Reaction of B to A's EAP		Description
Head		central, directed at A
Face	Forehead	crumpled
	Eyebrows	slightly raised & slightly drawn together
	Eyes	open & directed at A
	Mouth	slightly open
Torso		directed at the A
Hands		beside the body, holding hands, no movement
Body posture		aligned with the torso = directed at the A due to A's interfering, directly in the pathway of the pedestrian
Position		close to the A

Table 3: Sample nonverbal reaction of male native speaker of Icelandic (B) to the EAP of approaching female non-native speaker (A)

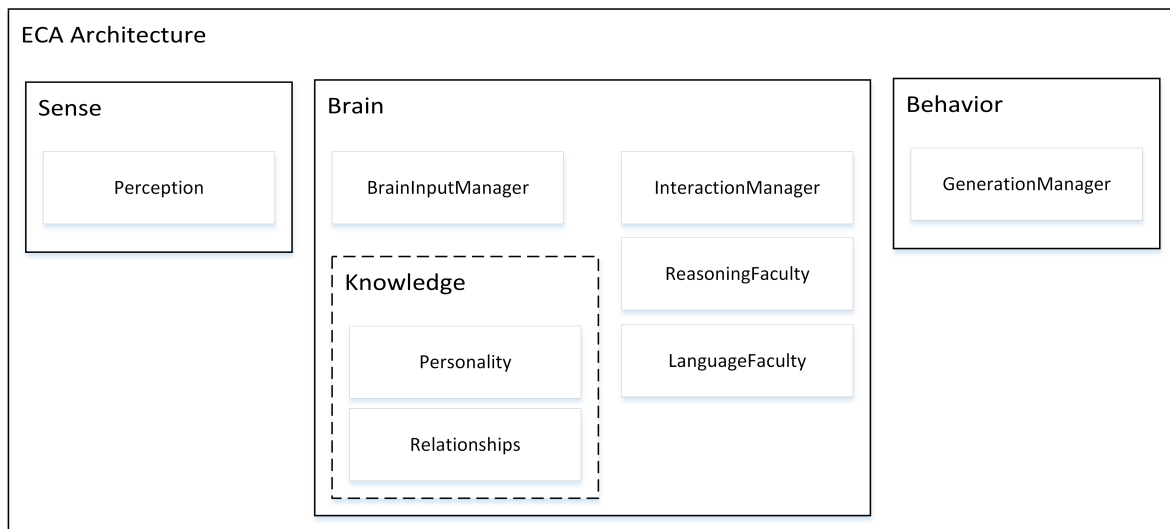


Figure 6: The Virtual Reykjavik ECA architecture. The Perception component acts as the agent's sensor and communicates with the BrainInputManager. The Brain's components work together with the discourse system in making communicative functions to be sent to the GenerationManager for behavior realization. [25]

called an FML document, which crucially includes the EAP. Following this, the other participant adds whatever functions he/she feels is necessary to the FML document, and finally the document is sent to each of the participants' behavior generation modules for processing. At this point, each ECA turns their respective FML information into BML (see section 4.1) and executes the relevant animations on the character.

Now the Approach block has reached an end state. The next time the floor calls for the next action to be executed, a new block must be selected. It is not predetermined which block that will be. It is important to note that if the two participants had known each other, they would have gone down a different path within the Approach block, entered into a greeting phase, and the EAP would never have happened.

6. Pilot User Study

A pilot user study has been conducted where six non-native speakers of Icelandic (four female, two male) played the first scenario in Virtual Reykjavik. Five of the subjects were beginners and one at intermediate level in Icelandic. In the context of the first task - to get someone's attention - the following words, although correct Icelandic, were used incorrectly in this context: *sæll* [how are you²] said to a male, *sæl* [how are you] said to a female, *blessaður* [how are you] said to a male person, *blessuð* [how are you] said to a female person, *halló* [hello], and *hæ* [hi]. The agents did not respond adequately because they were not designed for such greetings that are usually used among friends, acquaintances, and persons who know each other. These preliminary results indicate that the students in this sample were taught how to greet, but perhaps not how to approach a stranger on the streets as people do in real life. Further experimentation is needed to validate these findings.

The results also revealed that each user also used one of the three types of the EAP's verbal forms (see section 4.2) to announce his/her presence when approaching an ECA. In some cases, however, the users only approached the ECA and waited until it notices them. The proximity to the agent served the purpose of getting noticed. As it was a pilot study, in preparation for further testing, recordings of the computer screen were not made and therefore precise information on the proxemics was not retained, but will be included in the future.

7. Conclusion and Future Work

When teaching foreigners a new language, like Icelandic, it is imperative that they get lessons that reflect what happens in actual conversation. When analyzing situations where a stranger approaches another stranger, it became clear that the classic greeting phase [7] was missing. In Icelandic language lessons, foreigners are taught how to greet others [26]; however, this is not what we observed native speakers doing when non-native speakers, who were strangers, started conversations with them.

We observed that the EAP was the communicative function that most frequently occurred in situations where a stranger sought to initiate a conversation with another stranger for the purposes of asking for directions. This prompted the inclusion of such a function within the discourse models that arise during human-to-agent interaction. A model was implemented whereby the user EAP was the catalyst for conversation. Approaching an agent and clicking the mouse calls for an EAP, which prompts the user to speak and the conversation begins.

²There are not direct translations for these greetings, but they are forms not uttered between strangers

Early pilot tests have revealed that users may use inappropriate vocabulary when approaching native speakers in the simulated natural environment Virtual Reykjavik. This kind of vocabulary included greetings used among people who know each other and therefore not suitable for the EAP. Whatever the cause, future versions of the ECAs need to be aware of this tendency and be able to give the students constructive feedback. On the basis of our study, the EAP can potentially be generalized to other languages, because it seems to be a natural way how strangers approach other strangers in situations when they want to ask a question, e.g. directions to a particular place.

Acknowledgements

This work was made possible thanks to the Icelandic Research Fund (RANNÍS) and the collaborative spirit of researchers at the University of Iceland and the Center for Analysis and Design of Intelligent Agents (CADIA) at Reykjavik University.

8. References

- [1] B. Meyer, "Designing serious games for foreign language education in a global perspective," *Support for Learning*, vol. 1, pp. 715–719, 2009.
- [2] R. Ellis, *Task-based language learning and teaching*. Oxford University Press, 2003.
- [3] W. Littlewood, *Communicative language teaching: An introduction*. Cambridge University Press, 1981.
- [4] L. W. Johnson, H. Vilhjálmsson, and S. Marsella, "Serious games for language learning: How much game, how much AI?" *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, 2005.
- [5] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan, "Embodied conversational agents." Cambridge, MA, USA: MIT Press, 2000, ch. Human Conversation As a System Framework: Designing Embodied Conversational Agents, pp. 29–63.
- [6] J. Guðnason, O. Kjartansson, J. Jóhannsson, E. Carstensdóttir, H. Vilhjálmsson, H. Loftsson, S. Helgadóttir, K. Jóhannsdóttir, and E. Rögnvaldsson, "Almannaromur: An open icelandic speech corpus," in *Proceedings of the Third International Workshop on Spoken Language Technologies for Under-resourced languages (SLTU 2012)*, 2012.
- [7] A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press, 1990.
- [8] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The mumín coding scheme for the annotation of feedback, turn management and sequencing phenomena," *Language Resources and Evaluation*, vol. 41, no. 3/4, pp. 273–287, 2007.
- [9] F. Schiel, S. Steininger, and U. Türk, "The smartkom multimodal corpus at bas," in *LREC*, 2002.
- [10] K. Pápay, S. Szeghalmy, and I. Szekrényes, "Hucomtech multimodal corpus annotation," *Argumentum*, vol. 7, pp. 330–347, 2011.
- [11] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.
- [12] A. Cafaro, H. H. Vilhjálmsson, T. Bickmore, D. Heylen, and C. Pelachaud, "Representing communicative functions in saiba with a unified function markup language," in *Intelligent Virtual Agents*. Springer, 2014, pp. 81–94.
- [13] I. T. Coyne, "Sampling in qualitative research. purposeful and theoretical sampling; merging or clear boundaries?" *Journal of advanced nursing*, vol. 26, no. 3, pp. 623–630, 1997.
- [14] H. Bunt, J. Alexandersson, J. Carletta, J.-W. Choe, A. C. Fang, K. Hasida, V. Petukhova, A. Popescu-Belis, C. Soria, and D. Traum, "Language resource management—semantic annotation framework—part 2: Dialogue acts," *International Organization*, 2010.
- [15] I. Zwitterlood, A. Ozyurek, and P. M. Perniss, "Annotation of sign and gesture cross-linguistically," in *6th International Conference on Language Resources and Evaluation (LREC 2008)/3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*. ELDA, 2008, pp. 185–190.
- [16] H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud *et al.*, "The behavior markup language: Recent developments and challenges," in *Intelligent virtual agents*. Springer, 2007, pp. 99–111.
- [17] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari, "Multimodal human discourse: gesture and speech," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 9, no. 3, pp. 171–193, 2002.
- [18] S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, and H. Vilhjálmsson, "Towards a common framework for multimodal generation: The behavior markup language," in *Intelligent virtual agents*. Springer, 2006, pp. 205–217.
- [19] D. Heylen, S. Kopp, S. C. Marsella, C. Pelachaud, and H. Vilhjálmsson, "The next step towards a function markup language," in *Intelligent Virtual Agents*. Springer, 2008, pp. 270–280.
- [20] A. Cafaro, "First impressions in human-agent virtual encounters," 2014.
- [21] J. Allwood, L. Cerrato, L. Dybkjaer, K. Jokinen, C. Navarretta, and P. Paggio, "The mumín multimodal coding scheme," *NorFA yearbook*, vol. 2005, pp. 129–157, 2005.
- [22] S. Abrilian, L. Devillers, S. Buisine, and J.-C. Martin, "Emotv1: Annotation of real-life emotions for the specification of multimodal affective interfaces," in *HCI International*, 2005.
- [23] Á. Abuczki and E. B. Ghazaleh, "An overview of multimodal corpora, annotation tools and schemes," *Argumentum*, vol. 9, pp. 86–98, 2013.
- [24] H. Sloetjes and P. Wittenburg, "Annotation by category: Elan and iso dcr," in *LREC*, 2008.
- [25] S. Ólafsson, "When strangers meet: Collective construction of procedural conversation in embodied conversational agents," *Master's thesis, The University of Iceland, Reykjavik, Iceland*, 2015.
- [26] "Icelandic online," <http://icelandiconline.is/index.html>, accessed: 2016-15-01.