Exprgram: A Language Learning Interface for Mastering Pragmatic Competence

Kyung Je Jo
School of Computing, KAIST
Daejeon, Republic of Korea
kyungjejo@kaist.ac.kr

John Joon Young Chung
School of Computing, KAIST
Daejeon, Republic of Korea
johnr0@kaist.ac.kr

Juho Kim
School of Computing, KAIST
Daejeon, Republic of Korea
juhokim@kaist.ac.kr

Abstract

Mastering pragmatic competence, the ability to use language in a contextually appropriate way, is one of the most challenging parts of foreign language learning. Despite its importance, existing language learning systems often focus on linguistic components such as grammar, vocabulary, or pronunciation. Consequently, foreign language learners may generate grammatically flawless speech that is contextually inappropriate. With the diverse socio-cultural contexts captured in real-life settings, videos at scale can serve as a great material for learners to acquire pragmatic competence. We introduce Exprgram, a web-based video learning interface that assists learners to master pragmatic competence. With Exprgram, learners can raise their context-awareness, practice generating an alternative expression, and learn alternative expressions for the given context. Our user study with 12 advanced English learners shows potential in our learnersourcing approach to collecting descriptive context annotations and diverse alternative expressions.

Author Keywords

Learnersourcing; language learning; pragmatic competence; video learning

ACM Classification Keywords

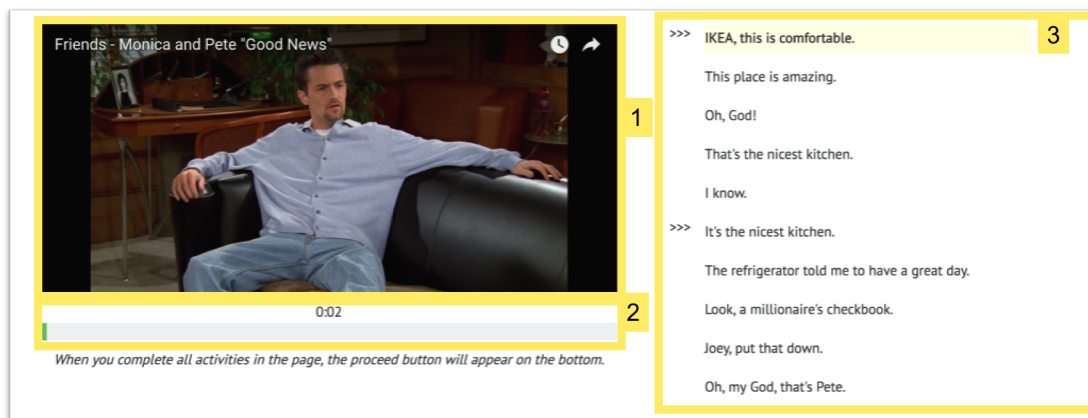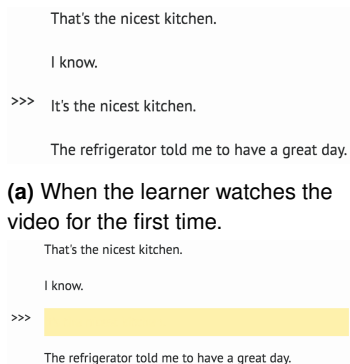H.5.m [Information interfaces and presentation (e.g.,HCI)]: Miscellaneous

Figure 1: Exprgram Interface. 1) Video player. 2) Unseekable progress bar for the system to control the playback. 3) Subtitles aligned with speech. The line of current utterance is highlighted in real-time.

(a) When the learner watches the video for the first time.

(b) In the second time watching the video, the target expression is masked and muted.

Figure 2: A target expression to learn is marked with >>>. An activity is prompted a few seconds after the video playback passes the target sentence.

Introduction

Pragmatic competence, the ability to use the language in a contextually appropriate way, is essential in foreign language learning [14]. Firth and Wagner assert that acquiring knowledge situated in physical and cultural context is key to foreign language learning [6]. In addition, exposing learners to as many real-life scenarios as possible is also important [8]. However, the limited and ideal scenarios in traditional learning materials provide narrow pragmatic opportunities to learners. Consequently, even proficient second-language learners may produce grammatically flawless speech that fails to achieve its communicative goal [1].

Foreign language videos naturally capture rich context such as settings, gestures, or emotions that can expose learners to the diverse usage of expressions [14]. The conversations in videos at scale can easily expose learners to diverse real-life contexts and expressions. For instance, in our analysis of 2,100 subtitles of English movies and dramas, "How are you?" appeared in 238 utterances, yielding diverse responses from commonly-taught (e.g., "good", "fine") to contextual (e.g., "depends","cold", "I'm not drunk") to sarcastic responses (e.g., "That's the 'hey' that means 'I need something'."). With the vast spectrum of contexts and diverse expressions, videos at scale present a pedagogical opportunity to facilitate the mastery of pragmatic competence.

While many language learning tools use videos, existing platforms (e.g., Viki, VoiceTube, FluentU) and tools often focus on teaching linguistic components such as vocabulary, grammar, or listening [10, 3]. Seiyuu-Seiyuu is a learning platform that directly focuses on pragmatic competence through a voice-driven system by allowing learners to practice their intonation, pausing, and pitch that are important to pragmatic competence [4]. While similar expressions that appear in different videos can teach learners diverse usage of expressions, existing systems often do not take advantage of pragmatically rich resources at scale.

To combat the challenge, we introduced an early version of Exprgram, a language learning interface that enables context- and expression-based browsing [7]. Ideally, users can browse videos that share a particular context or use expressions with semantic similarity. To enable such browsing, we collected context and expression annotations by adopting learnersourcing, the form of crowdsourcing that uses learners to generate useful artifacts for future learners while participating in a meaningful learning activity [9]. However, challenges of our earlier exploration remained as to 1) limited educational activities, 2) the workflow that is highly dependent on voluntary participation, and 3) lack of assistance for incapable learners.

In this paper, we introduce newly designed Exprgram that

Types of speech acts:
- ○ Representatives ⓘ
- ○ Commissives ⓘ
- ○ Directives ⓘ
- ○ Declarations ⓘ
- ○ Expressives ⓘ

Write a word that best describes the speaker's intention.

Intention: [   ]

Figure 3: Learners categorize the speech act and describe the intention of the expression.

Anger: ●————————
Disgust: ●————————
Fear: ●————————
Joy: ●————————
Surprise: ●————————
Sadness: ●————————
Emotion: [   ] ⓘ

Figure 4: Learners estimate the intensity of each emotion category with a slider whose value represents how strongly the emotion is captured in the expression. Learners also describe the emotion with a word of their choice.

provides two-stage learning to help learners acquire pragmatic competence. We designed activities following the insights from prior studies on pragmatic competence. Learners complete activities that are designed to have pedagogical value and the results can assist future learners. Through a user study with 12 participants, we show potential that our workflow can successfully generate artifacts that can be helpful for future learners in developing pragmatic competence.

System Overview

Learning language usage along the context is highly emphasized in acquiring pragmatic competence [1, 8, 14]. Whereas context that is explicit such as location or intonation can be easily captured, foreign language learners naturally lack the ability to interpret the underlying message that can only be understood with the surrounding context. For example, Bouton's study showed that even the advanced foreign language learners have difficulty in interpreting indirect expressions as in the following dialog - Sue: "How was your dinner last night?" Anne: "Well, the food was nicely presented." [2]. In this paper, we focus on teaching two types of context that are often implicit, and thus difficult: the intention and emotion of the speaker. In implementing Exprgram, we had three design goals inspired by interviews with two TESOL-certified instructors and prior studies on pragmatic competence [1, 2, 8, 11].

G1: Encourage learners to raise awareness of the intention and the emotion of the speaker in the video.

G2: Assist learners to master the usage of an expression in a particular context.

G3: Provide learners with diverse alternative expressions in the same context.

Exprgram integrates a web-based video player with subtitles aligned to the speech of the video (Figure 1). To meet these design goals, we use prompts to support two-staged learning: 1) raising context-awareness and 2) mastering the proper usage of diverse expressions. Learners watch the same video once for each stage to complete different activities presented by prompts. Exprgram marks each video with several *target expressions*, indicated with ">>>" in the subtitle section (Figure 1), that learners are expected to master. A difference between the two stages is that the target expression is masked (Figure 2(b)) and the video is muted at the time of the expression in the second stage. For each stage, prompts appear with an instruction of the following activity a few seconds after the video playback passes that target expression. Learners can replay the video as many times as they wish to complete the activities.

In the selection of target expressions, we focused on identifying expressions that are frequently used in real-life. Taking into account that expressions can exist in various form (e.g., "How are you?", "How do you do?", "How's it going?"), we adopted a machine learning algorithm, Sent2Vec, to calculate semantic similarity of each expression in our dataset [12]. Our dataset consists of 1-million expressions from 2,100 subtitles of English movies and dramas. We selected expressions that appear more than 5 times in our dataset to be the target expressions including the counts of expressions with a high semantic similarity of 95%.

Stage 1: raising context-awareness

Prior research showed that guidance and awareness-raising tasks help foreign language learners in capturing the implied context in conversations [11, 2]. Thus, we designed two context-awareness raising activities (Figure 3, 4) in Stage 1 to fulfill the design goal **G1**: to encourage learners to understand the speaker's intention and emotion.

Responses to Intention activity

Speech Acts	Most selected Representatives		Your choice Representatives
Top 3 intentions:	1. affirming	2. claiming	3. alleging
Your response:		describe	

Responses to Emotion activity

Type of emotions	Average	Your response
Anger	1	0
Disgust	0	0
Fear	0	0
Joy	54	60
Surprise	48	30
Sadness	0	0
Top 3 emotions:	1. amazed 2. excited 3. astonished	
Your response:	happy	

Figure 5: After completing activities to identify speaker's intention and emotion, learners can reflect on and evaluate from the statistics of learnersourced artifacts on speaker's intention and emotion.

☐ Just watch how I win.

☐ Would you be there?

☐ So what do you think?

☐ Can you come?

☐ Would you be next to me?

Figure 7: Five alternative expressions of "What do you think?" that are either learnersourced or computed by a machine learning algorithm. In the video, a man is persuading a woman to watch his boxing match.

Stage 1 includes four prompts including an instruction prompt at the beginning of each activity. In the second prompt, learners identify intention by selecting the type of speech act and describing the speaker's intention in a word (Figure 3). In this paper, we used Searle's 5 classifications of speech acts [13] which are widely used in studies on pragmatic competence [8]. The description for each speech act is provided within the interface as a tooltip. In the next task prompt, learners label the emotion of the speaker based on Ekman's six basic emotion [5] and a one-word description (Figure 4). The intention and emotion description tasks are designed to give learners an opportunity to process language on their own. After the two task prompts, Exprgram displays statistics of learnersourced artifacts (Figure 5) on

Based on the context, guess the appropriate expression on the muted part.
Here is the brief contextual information that may be helpful:

Most selected types of speech:		Expressives
Top 3 Intentions	affirming	claiming	expressing
Top 3 Emotions	amazed	impressed	eager

Average Emotion					
Anger	Disgust	Fear	Joy	Surprise	Sadness
1	0	0	65	35	0

Write an expression that is contextually appropriate.

Figure 6: Activity of generation an alternative expression.

which learners can evaluate and reflect on their responses. In addition, the statistic is an attempt to assisting incapable learners to become aware of the context. In this prompt, we also collect qualitative feedback on how well the learner-sourced artifacts describe the given context.

Stage 2: mastering the proper usage of expressions
The first prompt of Stage 2, presented in Figure 6, is designed to fulfill **G2**: aiding learners to master the usage of an expression. Learners can practice generating contextually appropriate expressions by referring to the learnersourced context labels on the top. Through practicing, learners are encouraged to master an expression that is adequate to the given context. The next prompt is designed to fulfill the design goal **G3**: to teach diverse alternative expressions. As shown in (Figure 7), 5 alternative expressions that are either learnersourced by previous learners or computed by the machine learning algorithm [12] are presented to learners. In this prompt, learners are instructed to choose expressions that are contextually appropriate. We designed the prompt in this way to control the quality of alternative expressions that may be erroneous and ensure learners to be aware of all alternatives provided.

User Study

We conducted a preliminary user study to verify our learnersourcing approach that leverages videos at scale for language learning.

Design & Methods
We recruited 12 advanced English as second language speakers on campus, aged between 20 to 26. The participants used Exprgram to watch a 5-minute clip from an episode of *Friends*, an American sitcom, with 8 target expressions. Each participant was given a 10-minute tutorial before beginning the task. A list of 100 frequent words used to describe emotions and intentions were provided to assist participants to find appropriate responses to the activities. Each participant received 10 USD for joining a 30-minute study session. For the expression generation task (Figure 6), participants were advised not to generate the exact expression in the video explicitly in the tutorial session and within the instruction in the prompt. Participants answered a short survey after they completed the task which primarily questioned how each component was helpful in understanding the context and the usage of expression.

Preliminary Results
Analysis of Learnersourced Artifact: Three native speakers independently worked on the same activity and resolved differences through a discussion to generate the ground truth. Comparing the speech act voted by the majority to the ground truth, the accuracy was 62.5%. To evaluate the accuracy of intention and emotion annotations described in a word, three native speakers independently scored each artifact and made consensus through a discussion. For each expression, 9.25 out of 12 intention annotations were correct with a standard deviation of 1.6 and 8.38 out of 12 emotion annotations were correct with a standard deviation of 2.1. Overall, participants performed quite well in describ-

ing both the intention (77%) and the emotion (70%). For the six scales of emotion, three native speakers generated ground truth about whether a type of emotion exists. After comparing the results, we found that participants were capable of discerning the existence of each emotion categories, with 86% of accuracy.

In the expression generation activity, 96 alternatives were collected in total from participants. Fourteen alternatives were excluded as they exactly matched the original expression of the video. All the remaining alternatives submitted by learners did not overlap. As a result, 10 alternatives on average were learnersourced for each target expression.

Qualitative Analysis on Learning Experience: From the survey given after the tasks, we analyzed the participants' experiences using Exprgram. First of all, learners expressed that statistics on learnersourced intention and emotion gave them an opportunity to reflect and evaluate their own responses. Furthermore, the learners explained the artifacts displayed on statistics (Figure 5 helped them to identify that they misunderstood. On the other hand, most learners felt the difference of artifacts from their responses were within their acceptable range. From the 5 alternative expressions Exprgram provides for each target expression (Figure 7), learners reported that they were able to find expressions that they were less familiar with. However, one learner stated that figure expressions were not helpful as they were too similar to her response.

Conclusion and Future Work

In this paper, we designed a language learning interface by utilizing conversations in videos to facilitate the mastery of pragmatic competence. We introduced a guided learnearsourcing workflow that enables learners to raise their own context-awareness and learn diverse expressions in the

same context while annotating the context and alternative expressions. In our user study, learners showed potential to generate context annotations, which describes the scene accurately, and expressions that are adequate to the given context. There are multiple directions for future work. We plan to study whether and how using Exprgram give a learning gain. We also plan to explore how learnersourced artifacts can effectively aid future learners. As Exprgram deals with intent and emotion annotations that are highly subjective, we wish to explore aggregation techniques to provide learners more meaningful information.

Acknowledgements

REFERENCES

1. Kathleen Bardovi-Harlig and Beverly S. Hartford. 1990. Congruence in Native and Nonnative Conversations: Status Balance in the Academic Advising Session *. *Language Learning* 40, 4 (1990), 467–501.

2. Lawrence F. Bouton. 1994. Conversational implicature in a second language: Learned slowly when not deliberately taught. *Journal of Pragmatics* 22, 2 (1994), 157–167.

3. Gabriel Culbertson, Solace Shen, Erik Andersen, and Malte F. Jung. 2017a. Have your Cake and Eat it Too: Foreign Language Learning with a Crowdsourced Video Captioning System. In *CSCW*.

4. Gabriel Culbertson, Solace Shen, Malte F. Jung, and Erik Andersen. 2017b. Facilitating Development of Pragmatic Competence through a Voice-driven Video Learning Interface. In *CHI*.

5. Paul Ekman, Wallace V. Friesen, Maureen O'Sullivan, Alfred Chan, I Diacoyanni-Tarlatzis, Kelsey Heider, Ronald Krause, W A LeCompte, Tom K Pitcairn, and Pio Enrico Ricci-Bitti. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology* 53 4 (1987), 712–7.

6. Alan Firth and Johannes Wagner. 1998. SLA Property: No Trespassing! *The Modern Language Journal* 82, 1 (1998), 91–94.

7. Kyung Je Jo, John Joon Young Chung, and Juho Kim. 2017. Exprgram: A Video-based Language Learning Interface Powered by Learnersourced Video Annotations. In *HCOMP Workshop*.

8. Gabriele Kasper. 1997. Can Pragmatic Competence be taught? (1997).

9. Juho Kim, Robert C. Miller, Krzysztof Z. Gajos, Scott R. Klemmer, Thomas P. Moran, and Maneesh Agrawala. 2015. Thesis: Learnersourcing: Improving Learning with Collective Learner Activity.

10. Geza Kovacs and Robert C. Miller. 2014. Smart Subtitles for Vocabulary Learning. In *CHI*.

11. Mikio Kubota. 1995. Teachability of Conversational Implicature to Japanese EFL Learners. 9 (01 1995).

12. Matteo Pagliardini, Prakhar Gupta, and Martin Jaggi. 2017. Unsupervised Learning of Sentence Embeddings using Compositional n-Gram Features. *arXiv* (2017).

13. J.R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

14. Gay N. Washburn. 2001. Using Situation Comedies for Pragmatic Language Teaching and Learning. *TESOL Journal* 10, 4 (2001), 21–26.