A Vietnamese Dialog Act Corpus Based on ISO Standard 24617-2

Thi-Lan Ngo[1,2], **Khac Linh Pham**[2]
[1]University of Information and Communication Technology (ICTU), Thainguyen Vietnam
[2]University of Engineering and Technology (UET-VNU), Hanoi Vietnam

Abstract

The voice-based human-machine interaction systems such as personal virtual assistants, chat-bots, and contact centers are becoming increasingly popular. In this trend, conversation mining research also is getting the attention of many researchers. Standardized data play an important role in conversation mining. In this paper, we present a new Vietnamese corpus annotated for dialog acts using the ISO standard 24617-2 (2012), for emotions using using Ekman's six primitives (1972), and for sentiment using the tags "positive", "negative" and "neutral". Emotion and sentiment are tagged at functional segment level. We show how the corpus is constructed and evaluated. This is the first Vietnamese dialog act corpus.

Keywords: Vietnamese corpus, dialog act corpus, ISO 24617–2 standard, conversational mining.

1. Introduction

In recent years, an extremely rapid progress in speech processing and recognition technology has led the emergence of voice-based human-machine interaction systems such as mobile virtual assistants, contact centers, and chatbots. These applications accommodate different purposes but they all need to able to understand the conversation while interacting with the user. Therefore, along with this development trend of human-machine interaction systems through natural language, conversation mining studies such as conversation structure analysis, conversation topic modeling, user intent understanding, and user emotion or satisfaction identification have also evolved and attracted the attention of many researchers. In these research, standardized dialog act corpora are the foundation. It is widely accepted that dialogue act annotation is very valuable in furthering understanding of interaction structure, and also in the design of artificial spoken or text dialogue (Wrede and Shriberg, 2003; Stolcke et al., 2006). There were several dialog act corpora available to the research community like as TRAINS (Traum, 1996), VERBMOBIL (Alexandersson et al., 1998), SWBD-DAMSL (Jurafsky et al., 1997), MRDA (Shriberg et al., 2004), AMI (McCowan et al., 2005) and so on. However, different corpora often apply a different scheme or modifying the existing scheme for dialog act annotation to serve task-specific needs. This creates a hardship in comparison of results and conclusions obtained when using different approaches due to a wide scatter of data in terms of the used annotation. Currently, ISO 24617-2 standard (ISO, 2012) is seem as "lingua franca" for dialog act annotation (Chowdhury et al., 2016; Bunt et al., 2012). Experimental studies in DBOX (Amanova et al., 2016) and DialogBank (Wijnhoven, 2016) corpora have shown good effects of the ISO standard on Dialog act annotation. Thus, in our work, we build a Vietnamese spoken corpus, the ViDa corpus, annotated dialog act according to the ISO 24617–2 standard (ISO, 2012), emotion tagging at functional segment (FS) level according to the Ekman's list of basic emotions (Ekman, 1972) and sentiment annotation at FS level. Addition, to make our purpose more useful in the intelligent systems using conversational interface and also for conversation mining purpose, we annotated meta information such as gender, dialect of users. The differences in our work compare to previous studies and the contribution is that:

- First, this is the first dialog act corpus for Vietnamese. Our corpus is not only annotated all labels, dimensions, relation defined in the ISO 24617–2 standard but also annotated meta information such as gender, dialect of participants.
- Second, this is the first corpus ever that annotate emotion, sentiment at functional segment level. All previous corpus only annotate labels at sentence, turn or document level.
- Third, we create a Vietnamese dialect dictionary for Vietnamese automatic dialect/accent detection in spoken conversation systems.

To build the corpus, we use IARPA Babel Vietnamese Language Pack IARPA-babel107b-v0.7 (IARPA-babel107b) (Andrus, Tony, et al, 2017). A brief description of IARPA–babel107b is presented in Subsection 2.1. The process of our corpus annotation is shown in Figure 1 and detailed in Section 2. Subsection 2.2 is about the segmentation of turns in the corpus into Functional segment. In Subsection 2.3, we talk about applying ISO 24617-2 to dialog act annotation. Emotion tagging is presented in Subsection 2.4.1. Sentiment tagging is described in Subsection 2.4.2. Finally is the conclusion part with the plan for future developments of our corpus.

2. Corpus and Annotation

2.1. Dataset pre–processing

We select transcripts of Vietnamese conversations obtained by an automatic speech recognition (ASR) in the IARPA (Intelligence Advanced Research Projects Activity) Babel program for data annotation. IARPA Data is published in IARPA-babel107b on LDC [1]. IARPA–babel107b contains about 201 hours of Vietnamese conversational and scripted telephone speech with corresponding transcripts. The data

[1] https://catalog.ldc.upenn.edu/LDC2017S01

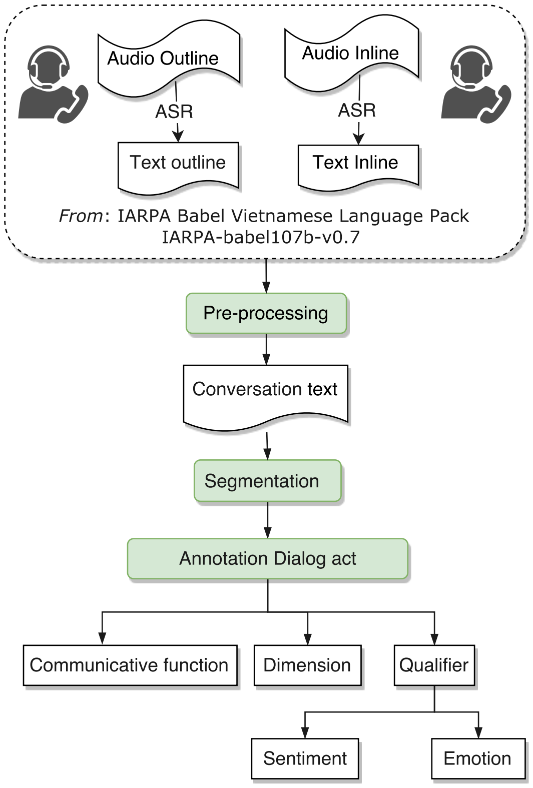

Figure 1: The process of data annotation.

is spoken in the North, Central and Southern dialect regions in Vietnam. We randomly select 28 dialogues in this dataset with any topic where number of dialog with each dialect regions is balance to build our dialog act corpus. Data in IARPA–babel107b is made in style in which a conversation between two persons includes a inline audio file with a inline text file (corresponding transcript of the inline audio file) and a outline audio file with a outline text file (corresponding transcript of the outline audio file). We make conversational texts from the inline texts and outline texts. After that, we review the conversation texts to rearrange it to the correct order of turns in the conversation using audio files. Error words from the results of ASR are retained. There are 1823 from error words in total 23803 words (92.1%). We note the meta information of our data including dialect regions and gender of participants, call time, duration of the phone calls, the number of turns in a conversation. In the pre–processing, we build a Vietnamese dialect dictionary includes 167 distinct southern Vietnam dialect words, 55 distinct central Vietnam dialect words and their translation to the ”standard” north Vietnam dialect. It is useful for automatic dialect/accent detection in spoken document retrieval systems. In human-machine interaction, it can help the system understand and communicate with users better. Instead of using the standard North Vietnamese dialect words for every machine, a friendly conversation interface application can detect the user’s dialect then use that dialect to communicate with the user.

2.2. Segmentation

Turns of conversation texts are segmented into functional segment (FS) unit, i.e., “minimal stretch of communicative behaviour that has one or more communicative functions” according to ISO 24617-2. Our corpus contain 28 dialogues, 2273 turns, 5065 functional segments. On average, each dialogue has 81.2 turns, 178.9 functional segments and each turn contain an average of 2.2 functional segments. The agreement scores of the segmentation process is 0.62 Fleiss kappa measure (Fleiss and Cohen, 1973).

2.3. Dialog Act Annotation according to ISO 24617-2

2.3.1. Dialog Act in ISO 24617-2

The ISO standard is amalgamated contributions from preexisting schemes, and is multifunctional and multidimensional - several communication acts can apply to stretches within the same contribution to the conversation of a participant. The ISO scheme see a dialogue act under 8 components, includes: (1) a sender; (2) one or more addressees; (3) a communicative function; (4) a semantic content; (5) a dimension; (6) functional dependence relations; (7) feedback dependence relations; and (8) rhetorical relations. In the dialog act annotation step, we annotated dimensions and dialog act for FSs. It contains 57 dialog acts in 9 dimensions: task, auto-feedback, allo-feedback, time management functions, turn management, discourse structuring, own communication own communication management, partner communication management, and social obligation management. The agreement scores of our Dialog Act annotation process is 0.76 Fleiss kappa measure.

Table 1: Distribution of dialog acts in the ViDa corpus

Dimension	**Number**	**Percent**
task	3137	60.72
autoFeedback	801	15.51
alloFeedback	19	0.37
turnManagement	533	10.32
timeManagement	353	6.83
discourseStructuring	186	3.60
ownCommunicationManagement	100	1.94
partnerCommunicationManagement	24	0.46
socialObligationsManagement	13	0.25

2.4. Qualifier

Our objective is to create an annotated corpus that will be a base resources for future researches in Vietnamese dialogue/conversation mining, namely such as suggestion mining, emotion mining, sentiment mining, request mining, argument mining. In this corpus, we label sentiment at functional segments level into 3 categories: positive, negative, neutral. The agreement scores of our sentiment annotation process is 0.85 Fleiss kappa measure.

Also in this corpus, we annotate emotions at functional segments level according to the Ekman’s (1972) list of basic emotions includes *joy, sadness, surprise, anger, fear* and *disgust*. We use *none* label for FS does not express emotion. There are many different taxonomy for labeling the

Table 2: Distribution of sentiment in the ViDa corpus

positive	489	9.76%
negative	655	13.07%
neutral	3866	77.17%

emotion, but we use Ekman's because it is used widely, is popular among researchers and simple enough to add into the ISO dialog act schema. The agreement scores of our emotion annotation process is 0.82 Fleiss kappa measure.

Table 3: The distribution of emotion in ViDa corpus

anger	205	4.09%
disgust	126	2.51%
fear	129	2.57%
joy	383	7.65%
sadness	313	6.25%
surprise	358	7.15%
none	3496	69.78%

Sentiment and emotion annotation at FS level has many advantages in sentiment and emotion analysis field compare to other levels. Previous studies in this field usually performed at the sentence/turn level or document level. Turn/sentence/document can be too long and contain more than one emotions or sentiments. Emotion/Sentiment annotation of FSs, the smallest part of sentence/turn that has the meaningful communicative function, will help us to understand emotions in turn/sentence/document more concretely. Also, because FS tend to be much shorter than turn/sentence/document, the sentiment and emotion classification at FS level can be much easier and be able to achieve higher precision.

3. Conclusion

Vietnamese is a very low-resource language. With the number of almost 100 millions speakers around the world (one of the most most spoken language) and the fast growing economy, the demand for Vietnamese standardized resource is greater than ever. Our corpus is aim to provide a first base resource for a variety of potential researches in Vietnamese natural language processing: mining suggestion, request, emotion, sentiment, argument in conversation, dialogue; Vietnamese dialog act identification; detection of user's gender, dialect. In the future, we intend to increase the size of our corpus and study deeper into the specific approaches of the these potential researches. We also intend to integrate them into real application such as personality virtual assistants, chat bots, contact center.

Vietnamese is a very low-resource language. With the number of almost 100 millions speakers around the world (one of the most most spoken language) and the fast growing economy, the demand for Vietnamese standardized resource is greater than ever. Our corpus is aim to provide a first base resource for a variety of potential researches in Vietnamese natural language processing: mining suggestion, request, emotion, sentiment, argument in conversation, dialogue; Vietnamese dialog act identification; detection of user's gender, dialect. In the future, we intend to increase the size of our corpus and study deeper into the specific approaches of the these potential researches. We also intend to integrate them into real application such as personality virtual assistants, chat bots, contact center.

Vietnamese is a very low-resource language. With the number of almost 100 millions speakers around the world (one of the most most spoken language) and the fast growing economy, the demand for Vietnamese standardized resource is greater than ever. Our corpus is aim to provide a first base resource for a variety of potential researches in Vietnamese natural language processing: mining suggestion, request, emotion, sentiment, argument in conversation, dialogue; Vietnamese dialog act identification; detection of user's gender, dialect. In the future, we intend to increase the size of our corpus and study deeper into the specific approaches of the these potential researches. We also intend to integrate them into real application such as personality virtual assistants, chat bots, contact center.

Appendix

4. References

Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Kipp, M., Koch, S., Maier, E., Reithinger, N., Schmitz, B., and Siegel, M. (1998). *Dialogue acts in Verbmobil 2*. DFKI Saarbrücken.

Amanova, D., Petukhova, V., and Klakow, D. (2016). Creating annotated dialogue resources: Cross-domain dialogue act classification. *Inform*, 26(11.5):36–0.

Andrus, Tony, et al. (2017). *IARPA Babel Vietnamese Language Pack IARPA-babel107b-v0.7 LDC2017S01*. IARPA (Intelligence Advanced Research Projects Activity) Babel program, distributed via LDC, ISLRN 401-277-958-467-7.

Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A. C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. R. (2012). Iso 24617-2: A semantically-based standard for dialogue annotation. In *LREC*, pages 430–437.

Chowdhury, S. A., Stepanov, E. A., and Riccardi, G. (2016). Transfer of corpus-specific dialogue act annotation to iso standard: Is it worth it? In *LREC*.

Ekman, P. (1972). Universal and cultural differences in facial expression of emotion. In *LREC'12*, page 207283. Nebraska Symposium on Motivation.

Table 4: Information Dialogue in ViDa corpus

STT	Dialogue	Participant 1		Participant 2		Number of Turns
		Dialect	*Gener*	*Dialect*	*Gener*	
1	D01_121544	North	female	Central	male	112
2	D02_013915	North	male	North	male	88
3	D03_222039	Central	female	Central	female	77
4	D04_002213	North	female	North	female	109
5	D05_165823	North	female	North	female	79
6	D06_200633	South	male	South	female	62
7	D07_225133	North	male	North	male	83
8	D08_162435	Central	female	Central	female	10
9	D09_203451	North	female	North	male	73
10	D10_202308	North	female	North	male	42
11	D13_183537	North	male	North	male	94
12	D14_014233	Central	male	Central	female	46
13	D15_182837	North	female	North	female	111
14	D16_202407	South	male	North	female	90
15	D17_023815	South	female	North	female	150
16	D18_160344	Central	male	Central	female	8
17	D19_162645	South	male	North	female	13
18	D20_151856	North	male	North	male	191
19	D22_005928	Central	female	Central	female	19
20	D24_130313	North	male	North	female	136
21	D26_120203	North	male	North	male	33
22	D27_223011	South	female	South	male	129
23	D28_003504	South	male	South	female	106
24	D29_134735	South	male	South	female	85
25	D39_173220	South	male	South	male	154
26	D47_192712	North	female	North	female	54
27	D51_002706	Central	male	Central	female	72
28	D53_010928	North	male	North	male	48

Fleiss, J. L. and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.

ISO. (2012). *ISO 24617–2: 2012 – Language resource management – Semantic annotation framework (SemAF) - Part 2: Dialog acts*. Geneva, Swizerland: International Organization for Standardization, ISLRN 401-277-958-467-7.

Jurafsky, D., Shriberg, E., and Biasca, D. (1997). Switchboard swbd-damsl labeling project coders manual, draft 13. *Technical Report 97–02*.

McCowan, I., Carletta, J., Kraaij, W., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., et al. (2005). The ami meeting corpus. In *International Conference on Methods and Techniques in Behavioral Research*, volume 88.

Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., and Carvey, H. (2004). The icsi meeting recorder dialog act (mrda) corpus. Technical report, INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. (2006). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Dialogue*, 26(3).

Traum, D. (1996). Conversational agency: The trains-93 dialogue manager. In *In Susann LuperFoy, Anton Nijhholt, and Gert Veldhuijzen van Zanten, editors, Proceedings of Twente Workshop on Language Technology, TWLT-II*. Citeseer.

Wijnhoven, K. (2016). Annotation representations and the construction of the dialogbank.

Wrede, B. and Shriberg, E. (2003). Relationship between dialogue acts and hot spots in meetings. In *Automatic Speech Recognition and Understanding, 2003. ASRU'03*, pages 180–185. IEEE.

Table 5: Example

Speaker	**Turn transcription**	**ID**	**FS**	**Dimension:function**
S	Alo chào anh ạ (Alo Hi you)	da1	alo	discourseStructuring: opening
		da2	chào anh ạ (Hi you)	socialObligationsManagement: initialGreeting
A	à chào em em dạo này khỏe không (ah hi you how are you)	da3, da4	à (ah)	autoFeedback: autoPositive (fe:da2) discourseStructuring: opening
		da5	chào em (hi you)	socialObligationsManagement:returnGreeting (fu:da2)
		da6	em dạo này khỏe không (are you fine)	task:propositionalQuestion
S	<laugh >em khỏe lắm tháng sau em ra sài gòn <breath > anh - em hứa sẽ ra thăm anh (<laugh >I'm good I'm going to Saigon next month <breath >you - I promise I will come to visit you)	da7, da8	<laugh >	turnManagement:turntake timeManagement:stalling «joy »
		da9	em khỏe lắm (I'm good)	task:answer (fu:da6)
		da10	tháng sau em ra sài gòn (I'm going to Saigon next month)	task:inform
		da11, da12	<breath >	turnManagement:turnKeep timeManagement:stalling
		da13	anh - (you -)	ownCommunicationManagement:retraction
		da14, da15	em hứa se ra thăm anh (I promise I will come to visit you)	ownCommunicationManagement:selfCorrection (fu:da13) task: promise