Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 236–244
Marseille, 11–16 May 2020
© European Language Resources Association (ELRA), licensed under CC-BY-NC

Crowdsourcing in the Development of a Multilingual FrameNet: A Case Study of Korean FrameNet

Younggyun Hahm[1], Youngbin Noh[1], Jiyoon Han[2], Taehwan Oh[2], Hyonsu Choe[2], Hansaem Kim[2], Key-Sun Choi[1]
[1]KAIST, [2]Yonsei University
{hahmyg, vincenoh, kschoi}@kaist.ac.kr, {clinamen35, ghksl0604, choehyonsu, khss}@yonsei.ac.kr

Abstract

Using current methods, the construction of multilingual resources in FrameNet is an expensive and complex task. While crowdsourcing is a viable alternative, it is difficult to include non-native English speakers in such efforts as they often have difficulty with English-based FrameNet tools. In this work, we investigated cross-lingual issues in crowdsourcing approaches for multilingual FrameNets, specifically in the context of the newly constructed Korean FrameNet. To accomplish this, we evaluated the effectiveness of various crowdsourcing settings whereby certain types of information are provided to workers, such as English definitions in FrameNet or translated definitions. We then evaluated whether the crowdsourced results accurately captured the meaning of frames both cross-culturally and cross-linguistically, and found that by allowing the crowd workers to make intuitive choices, they achieved a quality comparable to that of trained FrameNet experts (F1 > 0.75). The outcomes of this work are now publicly available as a new release of Korean FrameNet 1.1.

Keywords: FrameNet, Multilingual, Crowdsourcing

1. Introduction

FrameNet is a large linguistic resource in which the meaning of text is represented using Frame Semantics (frames) (Baker et al., 1998; Fillmore et al., 2003). An English language version of FrameNet was created by the FrameNet Project at the International Computer Science Institute[1] (ICSI), which defined English frames and assigned them to English vocabularies in order to construct lexical units (LUs) consisting of a word and frame pair. During annotation, the process of which was formalized by Ruppenhofer et al. (2006), annotators chose a proper frame (framing) for each target word and corresponding frame elements for the arguments. Since being developed, FrameNet has become a useful resource in a variety of fields, and researchers are now showing increasing interest in not only English FrameNet, but also in multilingual FrameNets for other languages (Baker et al., 2018; Borin et al., 2010; Meurs et al., 2008; sub, 2003; Burchardt et al., 2006; Yang et al., 2018; Virk and Prasad, 2018). However, despite this interest, the construction of multilingual FrameNets remains challenging due to the expense and complexity of annotating over raw sentences, including those in English.

As an alternative, crowdsourcing can be used to construct multilingual FrameNets. Several studies have investigated the viability of constructing FrameNets via crowdsourcing and found that annotations obtained via a crowd of 10 people are reliable when compared to annotations provided by experts (Fossati et al., 2013; Dumitrache et al., 2018). Similar studies have also been conducted for non-English languages. For example, Ohara et al. (2018) employed a Kyoto case frame dictionary when performing framing via crowdsourcing on Japanese sentences. In terms of multilingual FrameNets, while a few studies have explored cross-lingual issues for target-languages, such as the absence of a suitable English frame for target-language words (Candito et al., 2014; Pedersen et al., 2018; Ohara et al., 2018; Nešpore-Bērzkalne et al., 2018), to the best of our knowledge, no studies to date have evaluated the effectiveness of unifying the crowdsourcing settings that specify the information provided to workers. Instead, each crowdsourced approach has employed different settings to guide crowd workers during framing tasks, such as the definition and corresponding exemplar sentences for a particular frame. Also, no studies have evaluated the benefit or lack thereof of providing English ICSI frame definitions or additional semantic information (e.g., predicate-argument structure) to non-native English speaking crowd workers, nor have there been any studies that evaluated whether crowdsourced annotations effectively capture the ICSI meaning of English frames both cross-culturally and cross-linguistically. This is important because even for parallel sentences that are translations of each other, the corresponding frames may differ for cultural and/or linguistic reasons (Baker et al., 2018; Torrent et al., 2018). The focus of the current study was to address these issues in the context of Korean FrameNet (KFN) (Park et al., 2014; Kim et al., 2016), which was created using the *transferring approach* by translating existing FrameNets and using the original frame annotations as the original-language sentences. In this approach, parallel sentences have the same frame annotations for each aligned word. In this paper, the term *transferred annotation* is used to refer to the original frame annotation of translated words. It should be noted that some transferred annotations are not valid as they do not consider cross-lingual issues in translation, such as cases where a translation changes the meanings of words. For example, in KFN, while the general word *do* is annotated with the frame **Intentionally_act**, it is translated into the Korean word *yeon-ju-hada*, which means to 'play a musical instrument'. In this case, the transferred annotation was the frame **Intentionally_act**, which did not match the specific meaning of the word.

In preparation for our experiments, we evaluated the effi-

[1]http://framenet.icsi.berkeley.ed

cacy of certain crowdsourcing settings for Korean crowd workers that were non-native English speakers and were not familiar with FrameNet. This was important as Dumitrache et al. (2018) found that crowd workers did not always read, understand, and/or reflect the frame definitions in their framing task. We then evaluated the quality of crowdsourced annotations for different combinations of settings in which the information provided in the tasks varied (e.g., providing English frame definitions vs. translated frame definitions).

Second, we evaluated the extent to which the crowdsourced results captured the meaning of ICSI English frames. To accomplish this, we assembled a dataset containing 500 randomly selected annotations from KFN. Then, to address the problem of invalid transferred annotations as described above, we engaged trained FrameNet experts to annotate those 500 sentences with *expert annotations*. We then collected a total of 10,000 crowdsourced annotations, including the same 500 sentence. The crowdsourced annotations were then compared to both the transferred and expert annotations as if the crowdsourced annotations were more similar to the expert annotations than to the transferred annotations, it would confirm the cross-lingual superiority of the crowdsourced approach.

The contributions of this study are as follows:

1. We evaluated the effectiveness of various combinations of crowdsourced settings for non-native English speakers we engaged to construct a multilingual FrameNet. We found that providing only exemplar sentences for crowd workers allowed them to make intuitive choices and increased their quality and efficiency.

2. We evaluated whether crowdsourced annotations accurately captured the meaning of ICSI English frames both cross-culturally and cross-linguistically and found that the quality of the crowdsourced results were better than those of the transferred annotations. We also completed a case study on frame perspectives.

3. As a result of this work, a new release of KFN 1.1 is available [2] with 8,064 new annotations that were assessed as having a quality comparable to that provided by trained FrameNet experts (F1 $>$ 0.75).

The remainder of this paper is organized as follows. The crowdsourcing settings are introduced in Section 2. and their effectiveness is evaluated in Section 3.. The results of a comparison between the crowdsourced, transferred, and expert annotations are presented in Section 4. along with a case study. Some discussion is presented in Section 5.. Finally, some conclusions that can be drawn from this work are summarized in Section 6..

2. Crowdsourced Tasks

2.1. Overview

In preparation for the experiments, several crowdsourcing settings (M1 to M5) were defined and are described in detail in Subsection 2.2.. To evaluate the efficacy of each setting, five different groups of crowd workers, one for each setting, performed framing and argument annotation tasks for 1900 sentences and 100 KFN sentences per each setting. Each KFN sentence was then evaluated by comparing the transferred annotation and the corresponding expert annotation. A total of five crowdsourcing tasks were completed, involving a total of 500 KFN sentences. A flowchart of this process is shown in Figure 1.

2.2. Method

In this work, constructing the FrameNet via crowdsourcing was considered to be a *single step* task that included two subtasks (Fossati et al., 2013): framing, in which suitable frames are selected for a given target word in a sentence, and argument annotation, in which the span and role of arguments are selected for a given target word in a sentence. Ten crowd workers were assigned to each annotation task (Dumitrache et al., 2018), and in each task, the workers were assigned a target word in a sentence along with corresponding frame candidates and were then asked to select a suitable frame from among them. They then selected semantic roles for the arguments for the target word in the next step.

The objective in the first part of the testing was to answer the following questions: (1) for non-native English crowd workers, is there a qualitative difference in the results if original frame definition (i.e., written in English) and/or translated frame definitions (i.e., written in Korean) are provided to workers? (2) Is it beneficial to provide additional semantic annotations (e.g., predicate-argument structures) to workers? (3) Finally, which crowdsourcing task design was the most effective in building a multilingual FrameNet in terms of quality and efficiency?

To answer these questions, five different sets of information were provided to workers to assess which combination was the most effective for non-native English speakers. These five sets of information were defined as follows:

M1. Only provide the exemplar sentences for the frame.

M2. Provide both the exemplar sentences (M1) and original definition (English) of the frame.

M3. Provide both the exemplar sentences for the frame (M1) and predicate-argument information for the target word.

M4. Provide both the exemplar sentences (M1) and translated definition (Korean) of the frame.

M5. Provide the exemplar sentences (M1) and translated definition of the frame, and predicate-argument information of the target word.

Argument annotation tasks are tasks in which crowd workers are asked to choose the span of an argument and its semantic roles. To simplify the task, the Korean semantic role labeling (SRL) corpora was used as it has been annotated for spans of arguments for each target word in sentences. This allows workers to simply choose semantic roles for given arguments by selecting them from a

[2] https://github.com/machinereading/koreanframenet

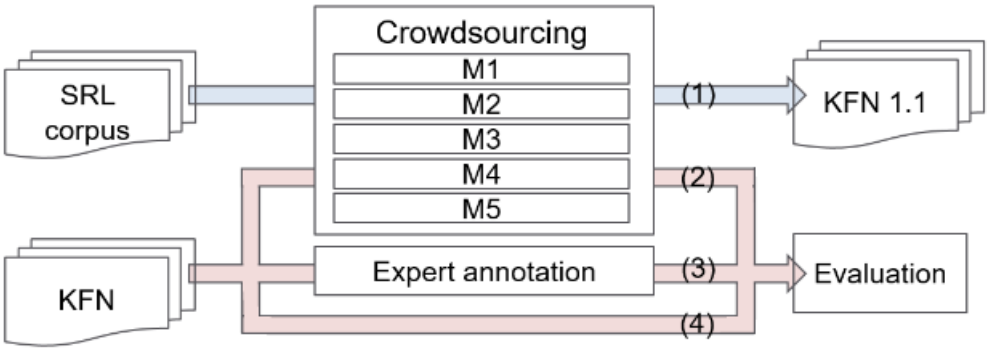

Figure 1: FrameNet annotation via crowdsourcing. (1) Crowd workers perform framing and argument annotation on a Korean semantic role labeling (SRL) corpus. (2) The quality of the crowdsourced annotations are evaluated by comparison with (3) expert and (4) transferred annotations, the latter of which were the original annotations in the existing KFN.

list. Propbank-style semantic annotations, such as ARG0, ARG1, and ARGM-LOC, were provided to workers as part of M3 and M5 in the form of predicate-argument structure information.

2.3. Dataset

The dataset used in this project was a subset of the Sejong dictionary (Kim, 2006) and a preprocessed dataset extracted from the Korean Propbank (Lee et al., 2015). The target words were verbs that had more than two frame candidates in the KFN LU database. To ensure only frequently used LUs were selected, the vocabulary was sorted based on the number of exemplar sentences in the original data, out of which a total of 2,370 words were selected. If there were more than 15 exemplar sentences for a particular word, only 15 sentences were used. In the end, a total of 9,500 sentences were collected, each of which had a target word, target word governing arguments, and Propbank-style semantic roles. FrameNet was referenced to obtain the original ICSI frame and semantic role definitions and the Korean frame definitions were translated via the Google translation service.

2.4. Crowdsourcing task

As mentioned previously, the crowdsourcing task consists of two subtasks, namely, framing and argument annotation. In the framing task, crowd workers chose suitable frames for a given target word from its frame candidates based on additional information, such as the frame definition and exemplar sentences. In the argument annotation task, crowd workers chose suitable semantic roles from those defined in FrameNet (i.e., frame elements) for each argument. Ten crowd workers were assigned to perform both subtasks in a single step, for which they were paid about $0.70 USD. The crowd workers have a bachelor's degree. The crowdsourcing annotation tasks were conducted five times using the different settings described in Subsection 2.2.. The dataset for each task consisted of 1900 sentences extracted from the crowdsourcing dataset and 100 sentences from KFN. In terms of evaluation, the crowd annotations were compared to the transferred annotations in KFN and the newly created expert annotations for same sentences.

2.5. Evaluation metric

The evaluation was conducted using CrowdTruth 2.0 [3], which provided annotation scores for each annotation created by the 10 crowd workers. In this tool, the scores were computed based on the ratio of workers that selected the frame to all workers that read the sentence. Then, the computed ratios were weighted by the quality of the worker. Here, the quality refers to the overall level of agreement between the results of one crowd worker and those of the other workers. Detailed evaluation metrics for this process are described in Dumitrache et al. (2018). The frame with the highest annotation score was considered to be the frame selected by the crowd workers for a particular sentence.
To evaluate the quality of the crowd annotations, the F1 score and accuracy were used in the comparison with the expert and transferred annotations. Thresholds were used to classify the annotation scores of the selected frames as either positive or negative.

2.6. Limitations

In terms of limitations, we evaluated only the framing tasks, but did not evaluate the argument annotation tasks. We have left this as an area of future work.

3. Evaluation 1: Effectiveness of Several Crowdsourced Task Designs

In this section, we provide the results of comparing crowdsourced annotations with expert annotations (in Figure 2). As part of the evaluation, different datasets were constructed for the five methods (M1 to M5) that each consisted of 1,900 newly created annotations and 100 annotations over KFN.

1) Is there a difference in the annotation quality between providing original frame definitions and translated frame definitions?
One of the main sources of error in crowdsourced annotations is that crowd workers often consider only exemplar sentences without frame definitions (Dumitrache et al., 2018), which may imply that ICSI English frame definitions are not helpful for non-native English speaking

[3]https://github.com/CrowdTruth/FrameDisambiguation

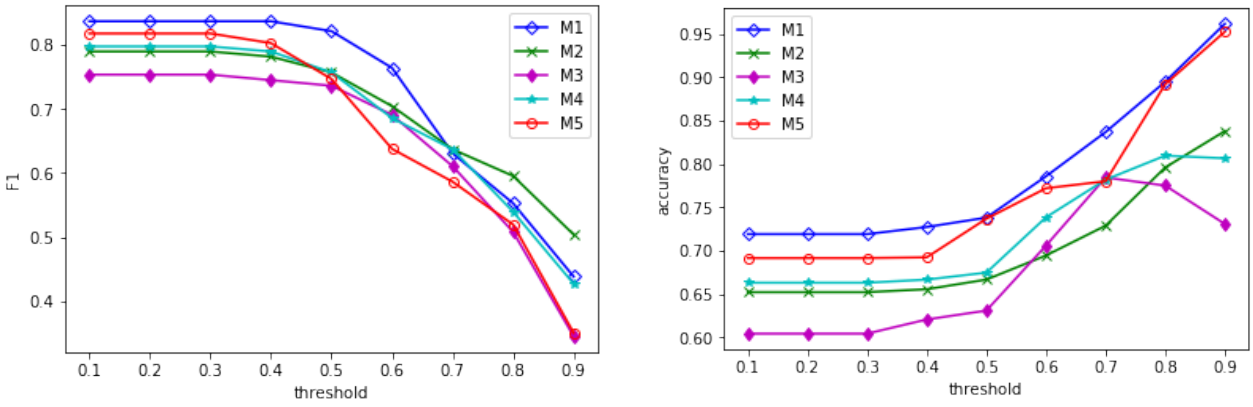

Figure 2: F1 score and accuracy for each crowdsourcing method. Each method employed exemplar sentences for the frame candidates (M1 to M5). In addition, M2 used the ICSI frame definitions (English), M3 uses the SRL annotations, M4 used the translated frame definitions (Korean), and M5 used both the translated frame definitions (Korean) and the SRL annotations.

crowd workers. Methods M2 and M4 were designed to test this hypothesis as M2 provided ICSI frame definitions (English) for crowd workers, while M4 provided translated frame definitions (Korean). When the results were evaluated, no significant difference was observed between the two groups of workers and the annotation quality. In practice, it seemed that some crowd workers had no difficulty understanding the ICSI frame definitions. In addition, many crowd workers ignored the frame definitions, but instead tended to choose frames based only on exemplar sentences and/or frame labels (e.g., the frame **Visiting** for the word *bang-mun-ha-da* 'visit').

2) Is providing other additional semantic annotations useful to crowd workers?

We initially assumed that providing additional semantic annotations would help crowd workers quickly gain an intuitive understanding of the meanings of words. To test this, only exemplar sentences were provided in M1 for each of the frame candidates, while exemplar sentences and Propbank-style predicate-argument annotations were provided in M3. As shown in Figure 2, the results from M3 were the lowest among all the methods. Moreover, the accuracy dropped even when the threshold was greater than 0.7.

This seems to be due to the difference in the semantic granularity between the FrameNet and Propbank annotations. In the Propbank-like annotations, the subject, objects, and modifiers are labeled as target words in a sentence; however, even though the words have the same arguments and syntactic and semantic roles in terms of Propbank, the words have different frames. Note that further studies are needed to better understand these results.

3) Which crowdsourcing task design is effective in building a multilingual FrameNet?

As shown in the Figure 2, in evaluating the F1 score, M1 achieved a score $>$ 0.83, while that for the others was $>$ 0.75. Even though they were trained in advance, it was evident that understanding the concept of FrameNet and frame definitions was not easy. Moreover, it was not easy to distinguish between two similar frames, even for experts (e.g., the frame **Certainty** vs. the frame **Likelihood**). In this case, providing additional information not only increased the workload, but also resulted in lower quality annotations. In contrast, methods M1 and M5 both achieved similarly good results. The results for methods M1 and M5 had an accuracy of $>$ 0.95 with a threshold of 0.8, while that for the others was $>$ 0.75. Roughly speaking, these results indicate that incomplete information caused the crowd workers to annotate differently and with lower quality.

Considering the costs, such as the time required to complete the task, a reasonable approach that is supported by these results is to only provide exemplar sentences for crowd workers, which would enable them to make intuitive choices and achieve higher quality and efficiency.

4. Evaluation 2: Quality of Crowdsourced Annotations

KFN was constructed via translation from the English and Japanese FrameNets to Korean. All existing KFN annotations were transferred annotations, which were the same as the original frames. In this section, we evaluate the quality of the crowdsourced annotations and compare the existing transferred frames with the *crowdsourced-frames* that were annotated via crowdsourcing.

4.1. Quality of the crowdsourced annotations

The results of a comparison between the crowd annotations and the expert and transferred annotations are shown in Figure 3. Note that the transferred annotations were obtained from the English and Japanese FrameNets, which rely on basic vocabularies and more high-level frames in general. In contrast, the crowd workers tended to select more specific frames instead of comprehensive frames, such as the frame **Intentionally_act**. As a result, the crowdsourcing approach better captured the meaning of the ICSI English frames both cross-culturally and cross-linguistically than those obtained via the transferring approach.

Data	Number of frames	Top 10 selected frames (occurrence)
Transferred annotations	249	Perception_active (13), Intentionally_act (12), Statement (11), Experiencer_focus (10), Possession (9), Motion (9), Attempt (8), Evidence (8), Leadership (7), Leadership (7), Perception_experience (6)
Crowdsourced annotations	223	Experiencer_focus (10), Intentionally_act (9), Statement (9), Creating (8), Possession (7), Participation (7), Perception_experience (6), Motion (6), Leadership (6), Emotion_active (5)

Table 1: Top 10 selected frames in the transferred annotations (in existing KFN) and crowdsourced annotations for the 500 sample sentences.

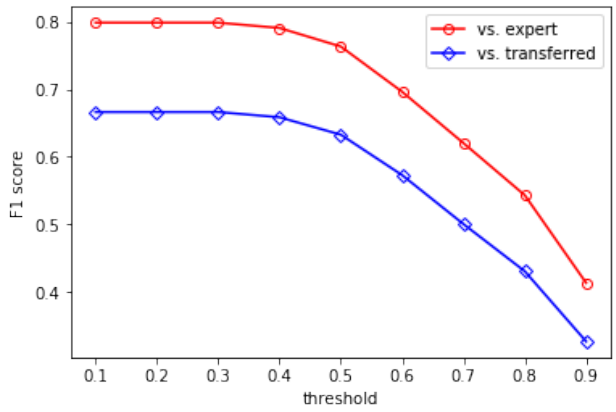

Figure 3: Comparing the crowdsourced annotations with the transferred and expert annotations.

The crowdsourced annotations were found to be of comparable quality to those provided by trained FrameNet experts (F1 $>$ 0.75) when the annotation scores as described in Subsection 2.5. were above a threshold of 0.5. In the results in Dumitrache et al. (2018), the quality of crowd annotation had an F1 score $>$ 0.67. The reason for the relatively high quality in that case was that all crowd workers were trained once an hour on the concept of FrameNet and the crowdsourced tasks. The 8,064 new annotations (F1 $>$ 0.75) produced in this work were compiled and are available in the new release of KFN 1.1.
A case study is described in the following subsections.

4.2. Relationship between the transferred and crowdsourced annotations

The characteristics describing the top 10 selected frames in the transferred annotations in the existing KFN and the crowd annotations on the 500 KFN sentences (described in Subsection 2.1.) are shown in Table 1. The ratio of the top 10 frames in the transferred annotations was 37.35%, and 32.74% in the crowd annotations.
The frame **Perception_active** was annotated 13 times in the original annotation; however, it was annotated only three times in the crowd annotations. The statistics of the other frequently occurring frames were also different. This suggests that there was a tendency for crowd workers to choose specific frames as good representations of the meanings of words rather than selecting higher concept frames. For example, the original word あじわう (*aziwau* 'taste, experience') was annotated with the frame **Perception_active**. However, it was translated into the word *meog-da* 'eat' because the word *aziwau* was used as a metaphor in the Japanese sentence to represent the concept of ingestion. The crowd workers chose the frame **Ingestion**.
Another example is the word *engage*, which is usually annotated with the frame **Intentionally_act** in FrameNet 1.7. An example is the sentence "... are less willing to ***engage*** in local benefit-seeking ..." In this case, the word *engage* is translated with the word *gae-ip-ha-da* 'be involved in'. The crowd workers tended to choose the frame **Participation** rather than the frame **Intentionally_act** because the word *gae-ip-ha-da* 'be involved in' was only used for the meaning of participation.
The case of the frame **Experiencer_focus** is more complex. A Korean word *jili-da* 'be frightened' was translated from several Japanese words, such as おびえる (*obieru* 'be frightened'), あきれる (*akireru* 'be surprised, be absurd'), and あきる (*akiru* 'tire of'). In some cases, the word *jili-da* 'be frightened' was annotated with the original frame **Experiencer_focus** when used in a sentence without objects or emotion. However, if there were objects and emotion in a sentence, the crowd workers generally selected the frame **Emotion_directed**.
Several other examples are listed in Table 2

4.3. Frame-frame relationships

In FrameNet, there are hierarchical frame-frame relationships between frames, such as *Inherited_by* (a child frame that inherits the meaning of a parent frame) and *Is_Causative_of* (a causality relationship between frames). We analyzed the frame-frame relationships between the crowdsourced and transferred frames in the existing KFN, and found that in the newly annotated 258 crowdsourced-frames, 42 frames (16.27%) had a frame-frame relationship with the original transferred frames. The occurrence of frame-frame relationships was: *Is_Inherited_by* (13), *Is_Causation_of* (8), *Inherits_from* (7), *Is_Used_by* (5), *Is_Perspectivized_in* (2), *Is_Preceeded_by* (2), *See_also* (2), *Uses* (2), and *Has_Subframes* (1).

(1) *neo-neun myeoch ga-ji mun-je-e* ***jun-bi-doe-eoss-na****?*
'You **ready** for some issues?'

As a result, the crowdsourced-frames were found to be more suitable for Korean words. In (1), the original target word *ready* was considered as the initial state before beginning an activity. Here, the transferred frame was **Activity_ready_state**. In contrast, the translated Korean verb

Transferred Frame	Crowdsourced Framing	Korean LUs
Perception_active (13)	Perception_active (3)	*chyeo-da-bo-da* 'look (at)', *mat-da* 'smell', *ba-ra-bo-da* 'look (at)'
	Distributed_position (1)	*gam-ssa-da* 'cover (up)'
	Intentionally_act (1)	*gat-da* 'have'
	Activity_ongoing (1)	*gye-sog-doe-da* 'continue'
	Attention (1)	*gwi-leul gi-ul-i-da* 'give attention to someone's story or opinion'
	Punctual_perception (1)	*dul-leo-bo-da* 'to look around'
	Perception_experience (1)	*deul-li-da* 'be heard'
	Evoking (1)	*tteo-ol-li-da* 'recall'
	Manipulation (1)	*man-ji-da* 'touch'
	Ingestion (1)	*meog-da* 'eat'
	Emotion_of_mental_activity (1)	*jeul-gi-da* 'enjoy'
Intentionally_act (12)	Intentionally_act (4)	*mat-da* 'take care of', *yeol-da* 'open', *yeon-ju-ha-da* 'play a musical instrument', *chi-da* 'hit'
	Possession (2)	*gaj-da* 'have'
	Participation (1)	*gae-ib-ha-da* 'intervene'
	Creation (1)	*nae-da* 'make'
	Event (1)	*yeol-li-da* 'be held'
	Assessing (1)	*pyeong-ga-ha-da* 'evaluate'
	Conduct (1)	*haeng-dong-ha-da* 'act'
	Discussion (1)	*hyeob-ui-ha-da* 'discuss'
Statement (11)	Statement (3)	*bal-eon-ha-da* 'speak', *seon-eon-ha-da* 'declare', *yeon-seol-ha-da* 'speech'
	Reporting (2)	*bo-do-ha-da* 'report', *bo-do-doe-da* 'be reported'
	Expressing_publicly (1)	*bal-pyo-ha-da* 'publish'
	Affirm_or_deny (1)	*bu-in-ha-da* 'deny'
	Cause_to_perceive (1)	*si-sa-ha-da* 'bespeak'
	Telling (1)	*jeon-ha-da* 'tell'
	Chatting (1)	*i-ya-gi-ha-da* 'conversation (with/about)'
	Reasoning (1)	*ju-jang-ha-da* 'claim'
Experiencer_focus (10)	Experiencer_focus (4)	*go-tong-bad-da* 'suffer', *nol-la-da* 'be surprised', *jin-jeong-ha-da* 'calm down', *hwa-na-da* 'get angry'
	Emotion_active (2)	*geog-jeong-ha-da* 'worry', *u-lyeo-ha-da* 'concern'
	Stimulus_focus (1)	*kkam-jjag-* 'be surprised (by)'
	Taking_sides (1)	*ban-dae-ha-da* 'oppose'
	Emotion_of_mental_activity (1)	*jeul-gi-da* 'enjoy'
	Emotion_directed (1)	*jil-li-da* 'be frightened'

Table 2: Examples of crowdsourced framing vs. transferred frames (Perceptional_active, Intentionally_act, Statement, Experiencer_focus) and the corresponding Korean LUs.

jun-bi-doe- 'prepare' was closer to an action for preparing another action rather than a state, so the crowdsourced-frame **Activity_prepare** was found to be more suitable. The frame-frame relationship was *Is_Preceded_by*.

(2) ... *jeom-lyeong-gwa tong-je-leul* ***kkeut-nae-go****, ... u-li-ui yo-cheong ...*
'... our request to **end** the occupation and the suffering ...'

In (2), the original word *end* was translated into the Korean word *kkeut-nae-go* 'finish'. The original frame was **Cause_to_end** and the crowdsourced-frame was **Process_end**. The frame-frame relationship was *Is_Causative_of*. In English FrameNet, a LU *end.v* is assigned with **Cause_to_end**, but not with **Process_end**. A LU *finish.v* is assigned with **Process_end**, but not with **Cause_to_end**. The translated Korean word *kkeut-nae-go*

'finish' is more similar to the meaning of *finish.v*.
These phenomena can be interpreted based on the understanding that English FrameNet was constructed from a basic vocabulary, and is therefore often annotated with comprehensive frames to better represent the meaning of each word. While there are many implicit frame-frame relationships (i.e., not defined in ICSI frames), these are not considered in this study

4.4. Case study on frame perspectives

Case 1. frame Perception_active
The frame **Perception_active** is evoked by the LUs representing the senses, such as *admire.v*, *attend.v*, *feel.v*, *gaze.v*, *glance.v*, *look,v*, *observe.v*, *smell.v*, and *view.v*. After the crowdsourced revision process, the frame **Perception_active** was newly annotated with 11 different frames. As described in Subsection 4.2., translated words have more specific meanings based on their context.

(3) ... *gunye honja-man-e yoenju-e* ***gui-rul guilyet-da***.
'... **listened** to her own performance ...'

In (3), the original target word *listened* was translated into the Korean idiom, *gwi-leul giul-i-da* 'give attention to someone's story or opinion'. However, as this idiom means "cautionary attitude," the crowdsourced-frame **Attention** is more suitable than the original transferred frame **Perception_active**.

(4) ... *se-gye yu-san-eul* ***dul-leo-bo-neun*** *yeo-haeng.*
'... travel looking around world cultural heritage.'
(J) 世界遺産を**見**て回る旅行

Here, the original Japanese target word 見 (*mi* 'to see') was annotated with the frame **Perception_active**; however, the phrase 見て回る (*mitemawaru* 'to look around'), was translated into the Korean word *dul-leo-bo-neun* 'to look around', although it could also be translated into "scan quickly." The crowdsourced-frame was **Punctual_perception**. In some cases, the frame **Punctual_perception** was used for temporal perception in a short time. For example, the English LU *glimpse.v* was annotated with the frame **Punctual_perception**, which is defined as "A PERCEIVER briefly perceives a PHENOMENON, often resulting in partial or uncertain perception." In general, it is common for the frame **Perception_active** to be used for an individual word in an idiomatic phrase. In such cases, the transferred annotation does not always provide a good representation of the meaning of the translated words.

Case 2. frame Intentionally_act
The frame **Intentionally_act** is a comprehensive high-level frame that covers a wide range of acts, and there are many LUs and derived frames associated with it. However, during our testing, the crowd workers tended to choose more specific frames that better represented the meaning of the words in sentences.

(5) ... *geu-deul-eun jag-eun ga-pan-dae-leul* ***seol-chi-ha-go*** ...
'They **set up** the small stand ...'

The Korean word *seolchi-hada* "to set up" is annotated with the frame **Intentionally_act** in existing KFN; however, the crowd workers did not understand this word as referring to the act of assembling and install a machine. Instead, it is considered to be a locating action referring to the specific place where a machine is installed. For this reason, in the crowd annotation, the frame **Placing** was annotated with the word *seolchi-hada* "to set up."

(6) ... *mi-sa-il gi-sul tong-je che-je-leul* ***hyeob-ui-ha-yeoss-eu-myeo*** ...
'... **engaged** in consultation with the MTCR ...'

An example is the word *hyeob-ui-ha-da* 'discuss (with)'. In (6), the phrase *engaged in consultation with* is translated into the Korean word *hyeob-ui-ha-* 'discuss (with)'. The crowd workers chose the frame **Discussion** for the word to represent the meaning of people discussing and cooperating to solve an issue. The word *mi-sa-il gi-sul tong-je che-je* 'MTCR' is considered as a core frame element TOPIC of the frame **Discussion** rather than a frame element ACT of the frame **Intentionally_act**.

Case 3. frame Statement
The frame **Statement** consists of noun and verb LUs that describe an act to send a message to some addressee, such as *explain.v*, *mention.v*, *describe.v*, and *address.v*. The core frame elements of the frame Statement are SPEAKER, TOPIC, MESSAGE, and MEDIUM, and the frame element ADDRESSEE is the non-core frame element. When given this frame, the crowd workers chose specific frames that represented a situation and context of communicative action, such as **Expressing_publicly** and **Reporting**.

(7) ... *beom-haeng seong-myeong-eul* ***bal-pyo-ha-ji*** *anh-ass-da..*
'... did not **announce** the murder statement for ...'

The word *bal-pyo-ha* 'announce' refers to both a speaking action and an official speech with purpose when it used in conjunction with the word *seong-myong* 'an (official) statement'. When assigned this frame, the crowd workers chose the frame **Expressing_publicly** for the word *bal-pyo-ha* 'announce' as it conveys the meaning of an official announcement.

(8) ... *geo-lae-e dae-hae-seo Sunday Telegraph-ga* ***bodo-ha-yeoss-da***.
'... The Sunday Telegraph **reported** on the ... transaction.'

Also, the word *bodo-ha-* '(a press) report' was annotated with the frame **Reporting** to emphasize the role of the speaker. In contrast, the crowd workers chose the frame **Statement** for the verbs *yeonsul-hada* 'speech', *jujang-hada* 'claim', and *suneon-hada* 'declare' because those verbs are not influenced by the role of the speaker.

Case 4. frame Experiencer_focus
The annotators suggested the frame **Emotions_of_mental_activity** or the frame **Emotion_active** for a sentence describing the emotion of an experience.

However, if a sentence was focused on an event or a situation that provoked emotions in an experience, the crowd workers tend to choose the frame **Stimulus_focus**.

(9) ... *o-laen si-gi-e geol-chyeo-seo* ***geog-jeong-haess-seub-ni-da***.
'... was **worried** over a long time.'

(10) *so-nyeon-i so-li-leul ji-leu-ja geu-deul-eun kkam-jjag* ***nol-lass-da***.
'When the boys yelled, they were **surprised**.'

In (9), the word *geog-jeong-haess-seub-ni-da* 'to worry' was annotated with the frame **Emotion_active** as it is more active in meaning. The frame **Stimulus_focus** was selected for a sentence describing the stimulus that provoked an emotion. In (10), an event *when the boys yelled* provoked an emotion *be surprised*. For the emotional frames, the crowd workers chose more specific frames and considered the context of sentences that focused on the experience or stimulus.

5. Discussions

1) Why are some transferred FrameNet annotations invalid?

Many multilingual FrameNet studies use the *transferring approach*, which utilizes English FrameNet as a seed and transfers it to the target-language FrameNet. Translating English lexical units (LUs) into the target-language by using a bilingual dictionary (Kanamaru et al., 2005; Borin et al., 2010) and/or a bilingual corpus (Park et al., 2014; Kim et al., 2016; Pedersen et al., 2018; Yang et al., 2018) would be considered as a starting point of target-language words, and its frame. Generally, these works are based on the assumption that the translated target-language words have the same frames as the original-language words. However, because of the differences in culture and linguistics (Baker et al., 2018), frames are shown in different manners in bilingual corpora (Torrent et al., 2018). More specifically, even if a pair of sentences in a bilingual corpus are semantically matched, each aligned word would have different semantics in terms of token-level annotation.

The first major issue in the transferred FrameNet annotation is that translation changes the meaning of words. More specifically, original target words are translated into more specific words. For example, as shown in the example (3) in Subsection 4.4., the original Japanese target word *listen* is translated into the Korean idiom, *gwi-leul gi-ul-i-da* 'give attention to'. Crowd workers selected the frame **Attention** rather than the transferred frame **Perception_active**

The second is the word alignment problem that occurs when a phrase is translated. For example, in the example (6) in Subsection 4.4., the phrase *engaged in consultation with* is translated into the Korean word *hyeob-ui-ha-da* 'discuss (with)'. In this case, the transferring approach assigns the original frame **Intentionally_act** of the individual word *engaged* to the translated word *hyeob-ui-ha-da* 'discuss (with)'. Crowd workers selected the frame **Discussion** to mean people discussing and cooperating to solve an issue.

Some transferred annotation is not valid by considering the general usage of the Korean word. In the example (7), in Subsection 4.4., the Korean word *bal-pyo-ha-da* 'announce' is translated from the word *announce*. The Korean word *bal-pyo-ha-da* 'announce' implies not only a speaking action (frame *Statement*) but also an official speech with purpose when it used with the word *seong-myong* 'an (official) statement'. Crowd workers selected the frame **Expressing_publicly** because it conveys the meaning of an official announcement.

However, in the framing task, the frame candidates for a given target word should be given for crowd workers, with exemplar sentences for each frame. Therefore, the crowdsourcing approach is more of a method of extending the existing FrameNet resources than a method of building the initial data. The transferring approach is still a worthwhile method for building initial FrameNet data as a starting point as described in this section. It should be noted that some incorrect annotations of KFN are because the meaning of frame is not fully considered during the manual translation process. Considering the meaning of frames in translation would ensure the quality of the transferred annotations.

2) Is the crowdsourcing approach also applicable to other low-resource languages?

In this study, we do not use any language-specific features and tools for dataset creation. Therefore, although this study is focused on the case study for Korean FrameNet, it can be applied to other languages as well.

6. Conclusions

The focus of this study was to evaluate crowdsourced approaches for constructing a multilingual FrameNet, specifically in the context of Korean FrameNet. First, we evaluated the effectiveness of crowdsourcing settings suitable for non-native English speaking workers. Second, we evaluated whether the crowdsourced annotations accurately captured the meaning of ICSI English frames. We discussed the effective crowdsourcing task design. We also showed cases where the crowdsourced annotations capture the meaning of ICSI English frames, cross-culturally and cross-linguistically, more effectively than the transferred annotations do. Upon analysis, we found that allowing crowd workers to make intuitive choices resulted in higher quality. In our findings, providing frame definitions caused the crowd workers to annotate differently and with lower quality. This may imply that some ICSI frame definitions would need to be revised for the target-language. We have left this as an area of future work.

The results of this work are now publicly available in a new release of KFN 1.1.

7. Acknowledgements

This work was supported by Institute of Information Communications Technology Planning Evaluation(IITP) grant funded by the Korea government(MSIT) (2013-0-00109, WiseKB: Big data based self-evolving knowledge base and reasoning platform)

8. Bibliographical References

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

Baker, C. F., Ellsworth, M., Petruck, M. R., and Swayamdipta, S. (2018). Frame semantics across languages: Towards a multilingual framenet. In *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts*, pages 9–12.

Borin, L., Dannélls, D., Forsberg, M., Gronostaj, M. T., and Kokkinakis, D. (2010). The past meets the present in Swedish framenet++. In *14th EURALEX international congress*, pages 269–281.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The SALSA corpus: a German corpus resource for lexical semantics. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May. European Language Resources Association (ELRA).

Candito, M., Amsili, P., Barque, L., Benamara, F., de Chalendar, G., Djemaa, M., Haas, P., Huyghe, R., Mathieu, Y. Y., Muller, P., Sagot, B., and Vieu, L. (2014). Developing a French FrameNet: Methodology and first results. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1372–1379, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Dumitrache, A., Aroyo, L., and Welty, C. (2018). Capturing ambiguity in crowdsourcing frame disambiguation. In *Proceedings of the Sixth AAAI Conference on Human Computation and Crowdsourcing, HCOMP 2018, Zürich, Switzerland, July 5-8, 2018.*, pages 12–20.

Fillmore, C. J., Johnson, C. R., and Petruck, M. R. (2003). Background to framenet. *International journal of lexicography*, 16(3):235–250.

Fossati, M., Giuliano, C., and Tonelli, S. (2013). Outsourcing framenet to the crowd. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 2, pages 742–747.

Kanamaru, T., Murata, M., Kuroda, K., and Isahara, H. (2005). Obtaining Japanese lexical units for semantic frames from Berkeley framenet using a bilingual corpus. In *Proceedings of the Sixth International Workshop on Linguistically Interpreted Corpora (LINC-2005)*.

Kim, J.-u., Hahm, Y., and Choi, K.-S. (2016). Korean FrameNet expansion based on projection of Japanese FrameNet. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 175–179, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Kim, H. (2006). Korean national corpus in the 21st century sejong project. In *Proceedings of the 13th NIJL International Symposium*, pages 49–54. National Institute for Japanese Language Tokyo.

Lee, C., Lim, S., and Kim, H. (2015). Korean semantic role labeling using structured svm. *Journal of KIISE*, 42(2):220–226.

Meurs, M.-J., Duvert, F., Bechet, F., Lefevre, F., and Mori, R. d. (2008). Semantic frame annotation on the French media corpus. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May.

Nešpore-Bērzkalne, G., Saulīte, B., and Grūzītis, N. (2018). Latvian framenet: Cross-lingual issues. In *Human Language Technologies–The Baltic Perspective: Proceedings of the Eighth International Conference Baltic HLT 2018*, volume 307, page 96. IOS Press.

Ohara, K., Kawahara, D., Sekine, S., and Inui, K. (2018). Linking Japanese framenet with Kyoto university case frames using crowdsourcing. In Tiago Timponi Torrent, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, May. European Language Resources Association (ELRA).

Park, J., Nam, S., Kim, Y., Hahm, Y., Hwang, D., and Choi, K.-S. (2014). Frame-semantic web: a case study for korean. In *International Semantic Web Conference (Posters & Demos)*, pages 257–260.

Pedersen, B., Nimb, S., Søgaard, A., Hartmann, M., and Olsen, S. (2018). A Danish FrameNet Lexicon and an Annotated Corpus Used for Training and Evaluating a Semantic Frame Classifier. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018.

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., and Scheffczyk, J. (2006). *FrameNet II: Extended theory and practice*. International Computer Science Institute, Berkeley, California. Distributed with the FrameNet data.

(2003). Surprise: Spanish FrameNet! In *Proceedings of the Workshop on Frame Semantics, XVII International Congress of Linguists (CIL)*. Matfyzpress.

Torrent, T. T., Ellsworth, M., Baker, C., and Matos, E. E. d. S. (2018). The multilingual framenet shared annotation task: a preliminary report. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, May. European Language Resources Association (ELRA).

Virk, S. M. and Prasad, K. V. S. (2018). Towards Hindi/Urdu framenets via the multilingual framenet. In Tiago Timponi Torrent, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, May. European Language Resources Association (ELRA).

Yang, T.-H., Huang, H.-H., Yen, A.-Z., and Chen, H.-H. (2018). Transfer of Frames from English FrameNet to Construct Chinese FrameNet: A Bilingual Corpus-Based Approach. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).