Smart Subtitles for Language Learning

Geza Kovacs

MIT CSAIL

gkovacs@mit.edu

Abstract

Language learners often use subtitled videos to help them learn the language. However, standard subtitles are suboptimal for vocabulary acquisition, as translations are nonliteral and made at the phrase level, making it hard to find connections between the subtitle text and the words in the video. This paper presents Smart Subtitles, which are interactive subtitles tailored towards vocabulary acquisition. They provide features such as vocabulary definitions on hover, and dialog-based video navigation. Our user study shows that Chinese learners learn over twice as much vocabulary with Smart Subtitles than with dual Chinese-English subtitles, with similar levels of comprehension and enjoyment.

Author Keywords

subtitles; interactive videos; language acquisition

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces – Graphical User Interfaces

Introduction

People often watch foreign-language videos with subtitles. In countries such as the Netherlands, for example, the majority of television programming is subtitled foreign-language content [1]. Language learning is a major reason why people watch foreign-language videos. For example, most Dutch viewers prefer subtitled foreign-language videos over dubbed videos, due to language-learning opportunities [1].

We aim to build a foreign-language video viewing tool that maximizes vocabulary learning while keeping the viewing experience enjoyable.

Background

Existing tools display various types of text on screen to help viewers comprehend foreign-language videos, which we define as follows:

- *Captions* show a transcript of the current line of dialog in the original language of the video. They are commonly used to assist the deaf and hard-of-hearing.

- *Native-language Subtitles* show a translation of the current line of dialog to the viewer's native language.

- *Dual Subtitles* show the native-language subtitle and the caption at the same time.

Figure 1 Dual subtitles shown using KMPlayer, with English subtitles on top, and a Chinese caption on the bottom.

Captions help learners with vocabulary learning [1] and pronunciation [2]. However, viewers who are not at an advanced level have difficulty comprehending and learning from authentic foreign-language videos with just captions, due to limited vocabulary knowledge [1].

Native-language subtitles aid comprehension, but detract attention from the foreign-language soundtrack [2]. They contribute to vocabulary learning, particularly if the learner makes efforts to associate unknown words in the dialog with the subtitles [1]. Dual subtitles make it easier to form this association by presenting both languages onscreen [3].

We hypothesize that since vocabulary comprehension is the main limitation of captions, then adding translation aids to captions would result in more vocabulary learning than dual subtitles.

Interface Features
We built an interactive video viewer geared towards learning by watching foreign-language videos. It has

Figure 2 Smart subtitles showing a word-level translation for the word the user has hovered over. Clicking on the 翻译 (translate) button will display the phrase-level translation.

features for video navigation and vocabulary acquisition, which we refer to as *Smart Subtitles*.

Dialog Display and Navigation
Smart Subtitles prominently display a transcript of the current line of dialog, as well as the preceding and following lines. We show the preceding line as a form of review, and the following line as a preview. Since learners often review and re-listen to individual lines of dialog, the navigation is dialog-based: clicking on a line of dialog will seek the video to the beginning of that utterance. The mouse wheel and arrow keys can also

be used for navigating the video by scrolling through the dialog lines.

Vocabulary Comprehension
Learners may encounter words they don't know while watching the video. With Smart Subtitles, they can obtain a definition in their native language by hovering over the word in the transcript. The best definition for that word in the context of the sentence will be displayed first, while alternative definitions are displayed afterwards, greyed-out to be less prominent. For languages such as Chinese or Japanese where the written form does not indicate the pronunciation, we also display the phonetic form above the transcript, in furigana for Japanese and pinyin for Chinese. Tones in the pinyin are represented with color in addition to diacritics, to make tones more visually salient to viewers; the tone colorization scheme is taken from the Chinese Through Tone and Color series.

While word-level definitions are enough for the learner to comprehend the video in many situations, some idiomatic expressions and grammar patterns require phrase-level translations. Thus, in addition to word-level translations, Smart Subtitles also provide a translate button that the user can click to display a translation of the current line of dialog, and pause the video so they have time to read it.

Implementation
Smart Subtitles are automatically generated from captions with the assistance of dictionaries and machine translation.

Obtaining Captions
Our system uses digital text captions in the WebVVT format [4] as input. We can download these from various online services, such as Universal Subtitles. We can also extract captions from DVDs, but because DVDs store captions as an image overlay instead of text, we first convert them to text via Optical Character Recognition (OCR), which may introduce errors.

Obtaining Word Definitions and Phrase Translations
The process of obtaining the words and their definitions depends on the language that is being learned. For Chinese and Japanese, which do not have spaces between words, we determine the words using statistical word segmentation [5]. For other languages, we split the text into words by spaces. Then, for each word we obtain the best definition for the word in the context of the sentence, as well as a list of alternative definitions. For Chinese, we get the best definition and pinyin from the Adsotrans software [6], and alternative definitions from CC-CEDICT [7]. For Japanese, we get definitions and furigana from WWWJDIC [8]. For other languages, we use Microsoft's Translation API.

To get phrase-level translations, we use native-language subtitles if they are available. Otherwise, we use Microsoft's Translation API.

User Study
We compared Chinese vocabulary acquisition between dual subtitles and Smart Subtitles using a within-subjects study. Our hypotheses were:

H1: Learners will learn more vocabulary when viewing videos with Smart Subtitles than with dual subtitles.

H2: Learners will enjoy watching the videos as much with Smart Subtitles as with dual subtitles.

Participants
We recruited 8 participants from intermediate level Chinese classes at MIT, who had actively studied Chinese for between 1.5 – 2.5 years. All had experience with watching Chinese videos with English subtitles.

Viewing Conditions
Each participant viewed a pair of 5-minute clips from the drama 我是老师 (*I Am Teacher*). We chose the clips because the content was conversational, and the vocab usage and pronunciations were standard Chinese.

Half of the participants saw the first clip with dual subtitles and the second with Smart Subtitles, while the other half saw the first clip with Smart Subtitles and the second with dual subtitles. For the dual subtitles condition we used KMPlayer, showing English subtitles on top and Chinese on the bottom. For the Smart Subtitles condition we used our software. The Chinese and English subtitles used in both conditions were extracted from a DVD and converted to text with OCR followed by manual corrections.

Before participants started watching each clip, we informed them that they would be given a vocabulary quiz afterwards, and that they should attempt to learn the vocabulary while watching the video. We also showed them how to use each video viewing tool, and told them they could watch the clip for as long as they needed, pausing and rewinding if desired.

Vocabulary Quiz
After the participant finished watching a video clip, we gave them an 18-question vocabulary quiz for the clip. The questions asked for English translations of Chinese

---

1) In the following sentence, what does the word 资格 mean?

我没有**资格**当老师

Meaning: _____
Did you already know the meaning of this word (资格) before watching this video?

Figure 3 Sample question from the quiz. Half of the questions were of this form; the other half did not provide a sentence.

words that had appeared in the video clip. For half of the questions, we also provided the sentence in which the word had appeared in the video as additional context. If the participant did not recognize the Chinese characters for the word, we provided them with pronunciations, and took note.

We excluded basic vocabulary words that are covered in beginner-level Chinese classes and words with no straightforward English translations from the vocabulary quiz. Additionally, to distinguish between words that were learned and those known beforehand, we asked participants to self-report whether they had already known the meanings of the words beforehand. If they claimed to have previously known the spoken form but not the written form of the word, we still considered them to have known the word beforehand.

Questionnaire
After participants completed the vocabulary quiz, they filled out a user-satisfaction questionnaire [9]. The questionnaire asked participants how easy they found it to learn vocabulary, how well they understood the video, and how enjoyable they found the viewing experience. It also asked them to summarize the video clip they had just seen, and provide feedback.
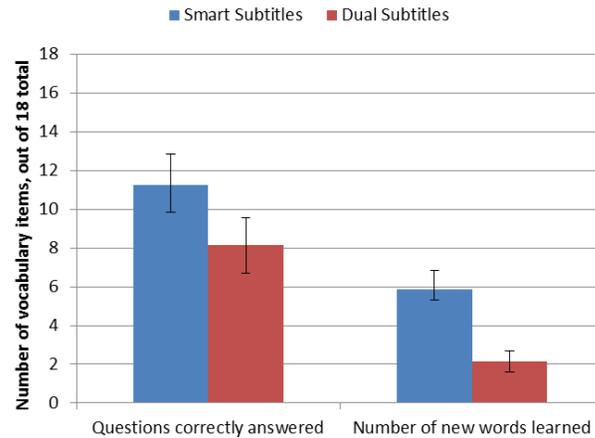
Figure 4 Vocabulary quiz results, with standard error bars
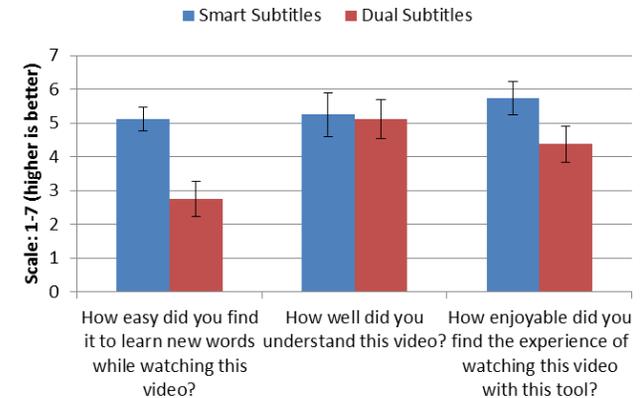
Results

Vocabulary Quiz Results

As illustrated in Figure 4, on average, participants reported in the quiz that they had already known 5-6 of the 18 vocabulary words before watching the clip, so we were testing about 12 new words in each 5-minute clip. There was no significant difference in the number of words known beforehand in each condition. A t-test shows that there were significantly more questions correctly answered (t=3.49, df=7, p < 0.05) and new words learned (t=5, df=7, p < 0.005) when using Smart Subtitles.

Providing pronunciations helped participants identify words they had previously known. However, participants did not require pronunciations for defining the meanings of most new words they learned in the video (on average, only 1 newly learned word was such

that they required pronunciation to recognize it). This suggests that they were actively reading the transcripts to learn how the words were written.

There were no significant differences in viewing times between the two tools. The average viewing time for a 5-minute clip with Smart Subtitles was 11.1 minutes (stddev=3.1), while the average viewing time with Dual Subtitles was 11.8 minutes (stddev=3.3).

Figure 5 Questionnaire results, with standard error bars.

Questionnaire Results

As illustrated in Figure 5, participants rated it easier to learn new words with Smart Subtitles, primarily attributing it to the presence of pinyin, and their ability to hover over words for definitions. A t-test showed that this difference was statistically significant (t=6.33, df=7, p < 0.0005). Participants rated their understanding of the clips as similarly high in both conditions.

Some participants considered the experience of watching using Smart Subtitles to be more enjoyable, attributing it to the overall interactive workflow: "I also really liked how the English translation isn't automatically there – I liked trying to guess the meaning based on what I know and looking up some vocab, and then checking it against the actual English translation".

Usage Observations
Most reviewing was local, with participants going back and pausing the video to read captions. Only 2 users re-watched the full clips with Dual Subtitles; none did so with Smart Subtitles. However, many of the participants re-read parts of the transcript after reaching the end of the clip when using Smart Subtitles.

Viewing strategies with Smart Subtitles varied across participants, though all made at least some use of both the word-level and phrase-level translation functionality. Word-level translations were heavily used - on average, users hovered over words in 3/4 of the lines of dialog. The words hovered over the longest tended to be less common words, indicating that participants were using the feature for defining unfamiliar words, as intended.

Participants tended to use phrase-level translations sparingly; on average they clicked on the translate button on only 1/3 of the lines of dialog. This suggests that word-level translations are often sufficient for learners to understand dialogs.

Conclusion and Future Work
We have presented Smart Subtitles, an interactive transcript with features to help learners, such as vocab definitions on hover and dialog-based video navigation.

Our user study found that participants learned more vocabulary with Smart Subtitles than dual Chinese-English subtitles, and rated their comprehension and enjoyment of the video as similarly high.

Although our study focused on vocabulary learning, the emphasis that Smart Subtitles place on the transcript and pinyin suggests that it may also be helpful for learning pronunciations for Chinese characters, and for learning sentence patterns.

While Smart Subtitles currently require users to actively interact with them, we could potentially allow more passive usage by predicting which words the viewer won't know and automatically showing their definitions.

References
[1] Danan, Martine. Captioning and Subtitling: Undervalued Language Learning Strategies. Translators' Journal, v49 (2004).
[2] Mitterer H, McQueen JM. Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception. PLoS ONE (2009).
[3] Raine, Paul. Incidental Learning of Vocabulary through Authentic Subtitled Videos. JALT (2012).
[4] WebVVT Standard. http://dev.w3.org/html5/webvtt
[5] Tseng, H. et al. A Conditional Random Field Word Segmenter. Fourth SIGHAN Workshop (2005).
[6] Adsotrans. http://www.adsotrans.com
[7] CC-CEDICT. http://cc-cedict.org
[8] WWWJDIC. http://wwwjdic.org
[9] Chen-Hsiang Yu and Robert Miller. Enhancing Web Page Readability for Non-native Readers. CHI (2010).