

Multimedia for Language Learning

by

Geza Kovacs

Submitted to the Department of Electrical and Computer Science on May 24, 2013 in partial fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology

May 2013

Copyright 2013 Geza Kovacs. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole and in part in any medium now known or hereafter created.

Author: _____
Department of Electrical Engineering and Computer Science

Certified By: _____
Robert C. Miller
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted By: _____
Prof. Dennis M. Freeman
Professor of Electrical Engineering and Computer Science
Chairman, Department Committee on Graduate Theses

Multimedia for Language Learning

by

Geza Kovacs

Submitted to the Department of Electrical and Computer Science on May 24, 2013 in partial fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science.

Abstract

Students studying foreign languages often wish to enjoy authentic foreign-language content - for example, foreign-language videos and comics. Existing means of presenting this content, however, are suboptimal from the perspective of language learning. We have developed a pair of tools that aim to help learners acquire the foreign language while enjoying authentic, foreign-language material. One tool is Smart Subtitles, which help learners learn vocabulary while watching videos. Our user evaluations have shown that Smart Subtitles help learners learn vocabulary more effectively than a popular approach for learning from video (dual subtitles). The other tool is a grammar visualization which illustrates the grammatical structure of sentences. This grammar visualization can be embedded into a number of language-learning applications. This includes a foreign-language manga reader we have built, which aims to help learners learn vocabulary, grammar and pronunciation while reading comics and manga. Our study for the grammar visualization shows that it helps monolinguals arrive at more accurate translations if we have an oracle taking the best translation of all, though there was no significant improvement in the average translation quality.

Thesis Supervisor: Robert C. Miller

Title: Professor

Table of Contents

Abstract	2
Table of Contents.....	3
Chapter 1: Introduction	5
Chapter 2: Related Work	9
2.1 Existing Ways to View Foreign-Language Videos	9
2.2 Custom Visualizations aimed towards Language Acquisition.....	11
2.3 Inserting words into Subtitles.....	13
2.4 Grammar Visualizations	14
2.5 Webpage Reading Assistance.....	16
Chapter 3: Smart Subtitles.....	20
3.1 Interface	20
3.1.1 Navigation Features	20
3.1.2 Vocabulary Learning Features	23
3.1.3 Excluded Features	26
3.2 Implementation.....	27
3.2.1 Obtaining Subtitles for Videos.....	27
3.2.2 Extracting Subtitles from Hard-Subtitled Videos.....	30
3.2.3 Getting Words, Definitions, and Romanizations	46
3.2.4 Translation-sense Disambiguation.....	47
3.3 User Study.....	51
3.3.1 Materials	51
3.3.2 Participants	51
3.3.3 Research Questions.....	51
3.3.4 Study Procedure.....	52
3.3.5 Results.....	53
Chapter 4: Grammar Visualization	60
4.1 User Interface.....	60
4.1.1 Applications: Manga Viewer.....	62

4.1.2 Applications: Reading Assistant.....	62
4.2 Implementation.....	63
4.2.1 Obtaining Phrase Structures	63
4.2.2 Phrase Translation	65
4.2.3 Voice Synthesis	66
4.3 User Study.....	66
4.3.1 Participants	66
4.3.2 Materials	66
4.3.3 Conditions.....	67
4.3.4 Procedure	68
4.3.5 Results.....	70
Chapter 5: Discussion.....	74
Chapter 6: Future Work and Conclusion.....	77
Bibliography	79
Appendices.....	82
Appendix 1: Questionnaire.....	82
Appendix 2: Sentences Used in Grammar Visualization Study	89

Chapter 1: Introduction

Students studying foreign languages often wish to enjoy authentic foreign-language content. For example, many students cite a desire to be able to watch anime and manga in its original form as their motivation for starting to study Japanese. However, the standard presentations of content are not accommodating towards language learners. For example, if a learner were watching anime, and did not recognize a word in the dialog, the learner's would normally have to listen carefully to the word, and look it up in a dictionary. This is a time-consuming process which detracts from the enjoyment of watching the content. Alternatively, the learner could simply watch or read a version that has already been translated to their native language to enjoy the content. However, they would not learn the foreign language this way.

We have produced a suite of tools to help people learn from authentic foreign language material. One is Smart Subtitles, shown in *Figure 1.1*, which is a video viewer aimed to help foreign-language learners learn vocabulary while watching foreign-language videos. It includes various features aimed at helping learners learn vocabulary from dialogs, and to navigate the video to help them review lines of dialog.



Figure 1.1: The interface for Smart Subtitles. As can be seen from the bubbles, the interface provides features for both dialog navigation as well as vocabulary learning.

This thesis also describes algorithms developed to support the Smart Subtitles system. This includes an algorithm to extract subtitle texts out of videos, as well as a translation-sense disambiguation algorithm to determine the best translation of a word in the context of a sentence.

Our user evaluation for Smart Subtitles consisted of having users watch a pair of Chinese-language video clips, one with Smart Subtitles, and another with dual Chinese-English subtitles. After each viewing, we administered a vocabulary quiz. We found that users learned more vocabulary using our system, with no decrease in perceived enjoyability or comprehension.

The other system is a visualization to help learners understand the grammatical structure of a sentence, shown in *Figure 1.2*. The general usage aims are to help language learners learn grammar, as well as to help them better comprehend sentences.

"Snake" is the general term for reptiles without feet

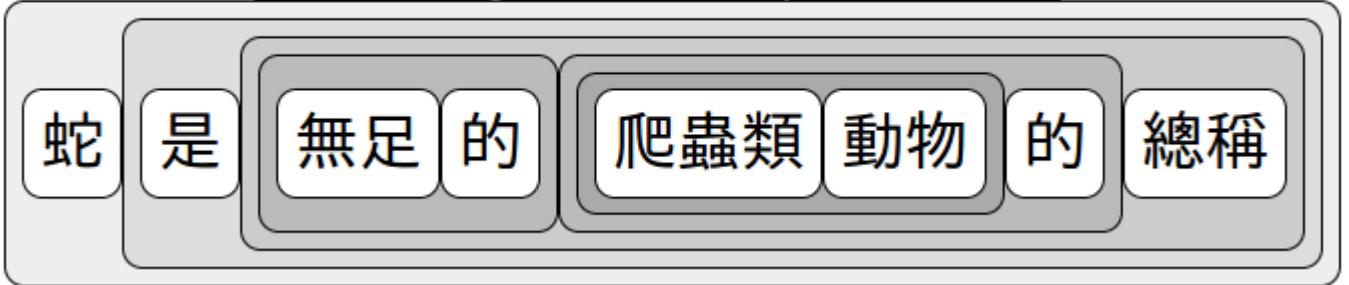


Figure 1.2: The grammar visualization displayed for a sentence.

This visualization can be embedded into a number of language-learning applications. This includes a foreign-language manga reader we have built, which aims to help people learn vocabulary, grammar, and pronunciation while reading comics and manga.



Figure 1.3: Our foreign-language manga viewer.

Another application we embedded our grammar visualization into is a foreign-language webpage reading assistant, which is a browser extension that allows users to click on any foreign-language sentence on a webpage, and have the grammar visualization appear, as seen in Figure 1.4.



Figure 1.4: Grammar visualization browser extension shows users the grammar visualization upon clicking on a foreign-language sentence in the webpage.

The evaluation of our grammar visualization focused on its ability to help novices (Mechanical Turk users who did not know Chinese, and whose native language was English) improve their comprehension of foreign-language sentences, relative to simply giving them a machine translation. Here, we measured comprehension by the quality of the translation that they are able to produce - thus making this a post-editing task. We observed a slight improvement in average translation quality when using our grammar visualization, relative to simply asking the users to correct the raw machine translation, or presenting them a visualization without the structural information, though the difference was not statistically significant. That said, if we take the highest-quality translation produced for each sentence in each category, then there is a significant improvement over the standard postediting condition. This suggests that our visualization can potentially improve users' comprehension of sentences.

This work's primary contributions show that:

- We can augment multimedia that is tailored towards entertainment purposes to make it educational material suitable for intermediate-level language learners.
- Our grammatical structure visualization can potentially scaffold even complete novices into comprehending and learning from foreign-language material.

Chapter 2: Related Work

2.1 Existing Ways to View Foreign-Language Videos

Videos in foreign languages are adapted for foreign audiences in many ways, which are illustrated in *Figure 2.1*:

One approach is *dubbing*, where the original foreign-language voice track is removed, and is replaced with a voice track in the viewer's native language. Because the foreign language is no longer present in the dubbed version, this medium is ineffective for foreign language learning.

Another approach is to provide *subtitles* with the video. In this approach, the foreign-language audio is retained as-is, and the native-language translation is provided in textual format, generally as a line presented at the bottom of the screen. Thus, the learner will hear the foreign language, but will not see its written form - therefore, they will need to pay attention to the audio. Subtitles have had mixed reactions in the context of language learning - some studies have found them to be beneficial for vocabulary acquisition, compared to watching videos without them [1]. That said, other studies have found them to provide no benefit to language learners in learning vocabulary [4]. Additionally, the presence of subtitles are considered to detract attention from the foreign-language audio and pronunciation [2]. The mixed results that studies have found on the effects of subtitles on language learning suggests that their effectiveness depends on factors such as the experience level of the learners [6].

In addition to subtitling and dubbing, which are the most commonly used in the context of foreign media consumption for enjoyment, there are some additional presentation forms that have been experimented with in the context of language acquisition:

With a *transcript* (also referred to as a *caption*), the video is shown along with the text in the language of the audio (in this case, the foreign language). Transcripts are generally used to assist hearing-impaired viewers; however, they can also be beneficial to language learners for comprehension, particularly if they have better reading ability in the foreign language than listening comprehension ability [1]. However, transcripts are only beneficial to more advanced learners whose language competence is already near the level of the video [6].

With *reverse subtitles*, the video has an audio track and a single subtitle, just as with regular subtitles. However, the key distinction is that this time, the audio is in the native language, and the subtitle shows the foreign language. This takes advantage of the fact that subtitle reading is a semi-automatic behavior [5], meaning that the presence of text on the screen tends to attract people's eyes to it, causing them to read it. Therefore, this should attract attention to the foreign-language text. The presentation of the foreign language in written form may also be helpful with certain learners whose reading comprehension ability is stronger than their listening comprehension. That said, because the foreign language is presented only in written form, the learner may not end up learning the pronunciation, particularly with a language with a non-phonetic writing system, such as Chinese.

With *dual subtitles*, the audio track for the video is kept as the original, foreign language. However, in addition to the subtitle displaying the foreign-language, they also display the viewer's native language as well. In this way, a learner can both read the written representation, as well as hear the spoken representation of the dialog, and will still have the translation available. Thus, of these options, dual subtitles provide the most information to the learner. Indeed, dual subtitles have been found to be at least as effective for vocabulary acquisition as either captions or subtitles alone [3]

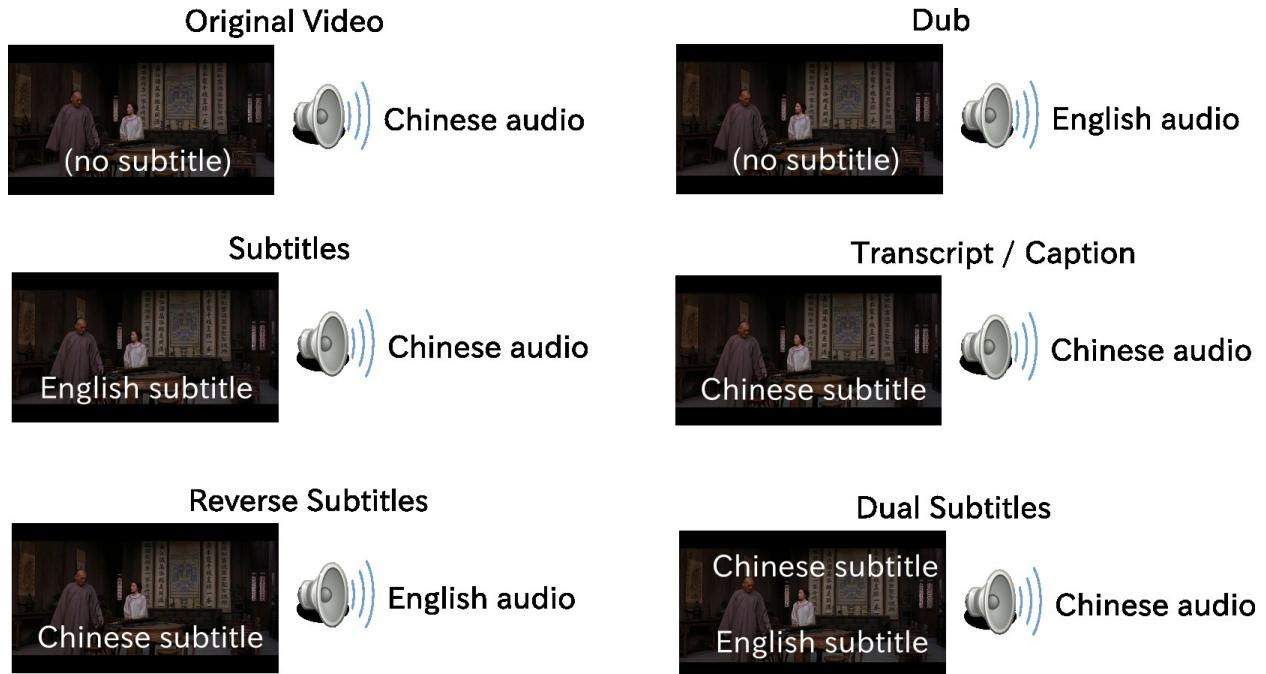


Figure 2.1: An overview of the existing ways one might present a Chinese-language video to learners whose native language is English.

2.2 Custom Visualizations aimed towards Language Acquisition

The interfaces we previously listed are general-purpose: given a video that has audio and subtitle streams in the appropriate languages, one can produce any of these subtitling conditions using an appropriate video player software, such as KMPlayer [7].

There are additionally other visualizations which are designed to help language learners enjoy multimedia. However, these currently have to be generated in a time-consuming manner using video editing tools.

For example, karaoke videos often highlight the word that is currently being said, and display a romanization for languages such as Chinese which do not use the Latin script, as shown in *Figure 2.2*.



Figure 2.2: A translation, romanizations, and word-level karaoke highlights are presented in this video to assist non-native speakers. [11]

Other instructional videos provide word-level translations, as shown in *Figure 2.3*:



Figure 2.3: Word-level translations are presented in this instructional video for Chinese learners [10]

Both the word-level translations and romanized pronunciations shown in these example instructional videos provided inspiration for developing our own video-viewing interface.

2.3 Inserting words into Subtitles

GliFlix [30], shown in *Figure 2.4*, augments video subtitles for language learning. It begins with subtitles, and inserts translations to the foreign language for the most common words that appear in the dialog. They attain larger rates of vocabulary acquisition compared to regular subtitles (though not dual subtitles) in user studies. Because GliFlix presents the foreign vocabulary in the order of the viewer's native language, this approach is likely ineffective for other language-learning tasks such as learning pronunciation and grammar. The system we will present takes the opposite approach to GliFlix. Instead of starting with a subtitle and inserting foreign-language vocabulary into it, we start with a transcript and attempt to help people understand it. We believe that our approach is preferable, not only because it focuses attention on the foreign language, but also because it is usable without a human-generated translation to English.



Figure 2.4: GliFlix showing in-subtitle translations for words.

2.4 Grammar Visualizations

The Chinese Room [21], shown in *Figure 2.5*, is targeted towards helping monolinguals comprehend machine translations by showing a word-by-word alignment between the machine translation and the original sentence. It visualizes sentences by showing a word-by-word alignment between the machine translation and the original sentence. However, this system has the disadvantage that visualizing a single sentence consumes the entire screen. Therefore, it is not suitable for usage as a drop-in replacement for the original foreign-language text, as it requires a context switch into a dedicated window to visualize an individual sentence. Many of its features, such as its visualization of the constituent parse tree, also require training and a background in NLP to comprehend. Indeed, they found that users with NLP experience performed better than ordinary users with their tool. Nevertheless, their user study found that users who had been trained to use the system demonstrated a gain in average translation quality relative to a baseline of simply post-editing a machine translation, at the expense of more than doubling edit times. However, the quality of the generated translations still lagged significantly behind translations generated by bilinguals.

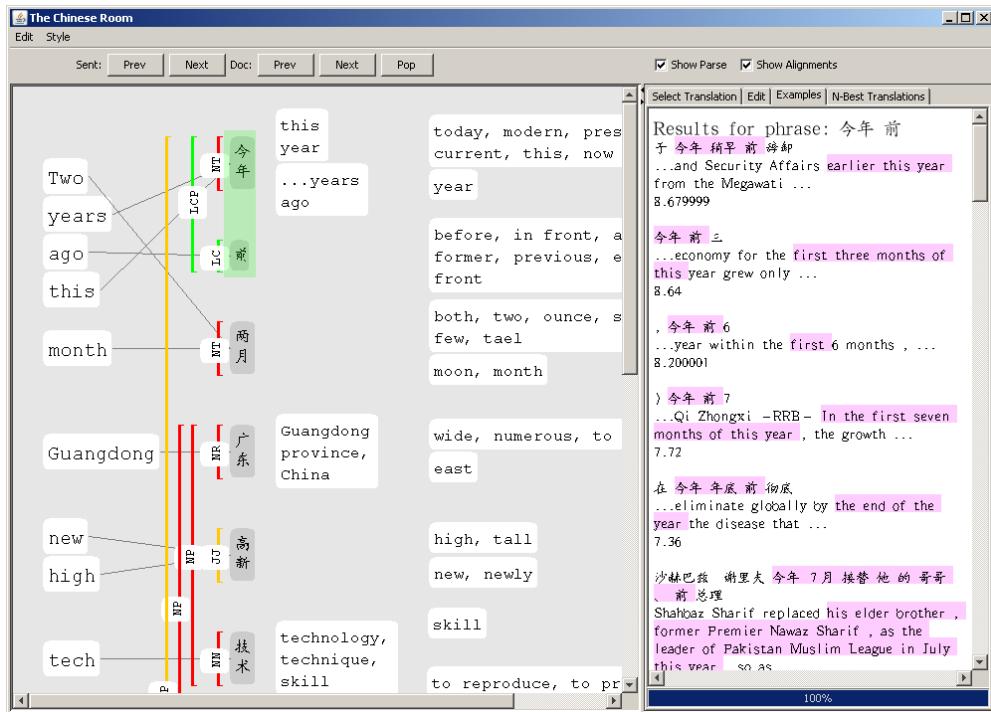


Figure 2.5: The Chinese Room visualizing a sentence.

DerivTool [28], shown in *Figure 2.6*, is another visualization for helping people generate translations. Specifically, it is an interactive rule-based machine translation system that prompts

users to select one of a number of possible translations for a given phrase, and uses this information to generate a machine translation. Because users are not actually generating the translation themselves, but are rather supplying information to the machine-translation system, this is not a general-purpose post-editing system. Additionally, because the system is based around applying syntactic rules to translate sentences, it can only be used with rule-based machine translation systems, whereas most state-of-the-art systems, such as Google or Microsoft's, are statistical machine translation system. Unfortunately, no user study was performed on the quality of translations generated by users using this system.

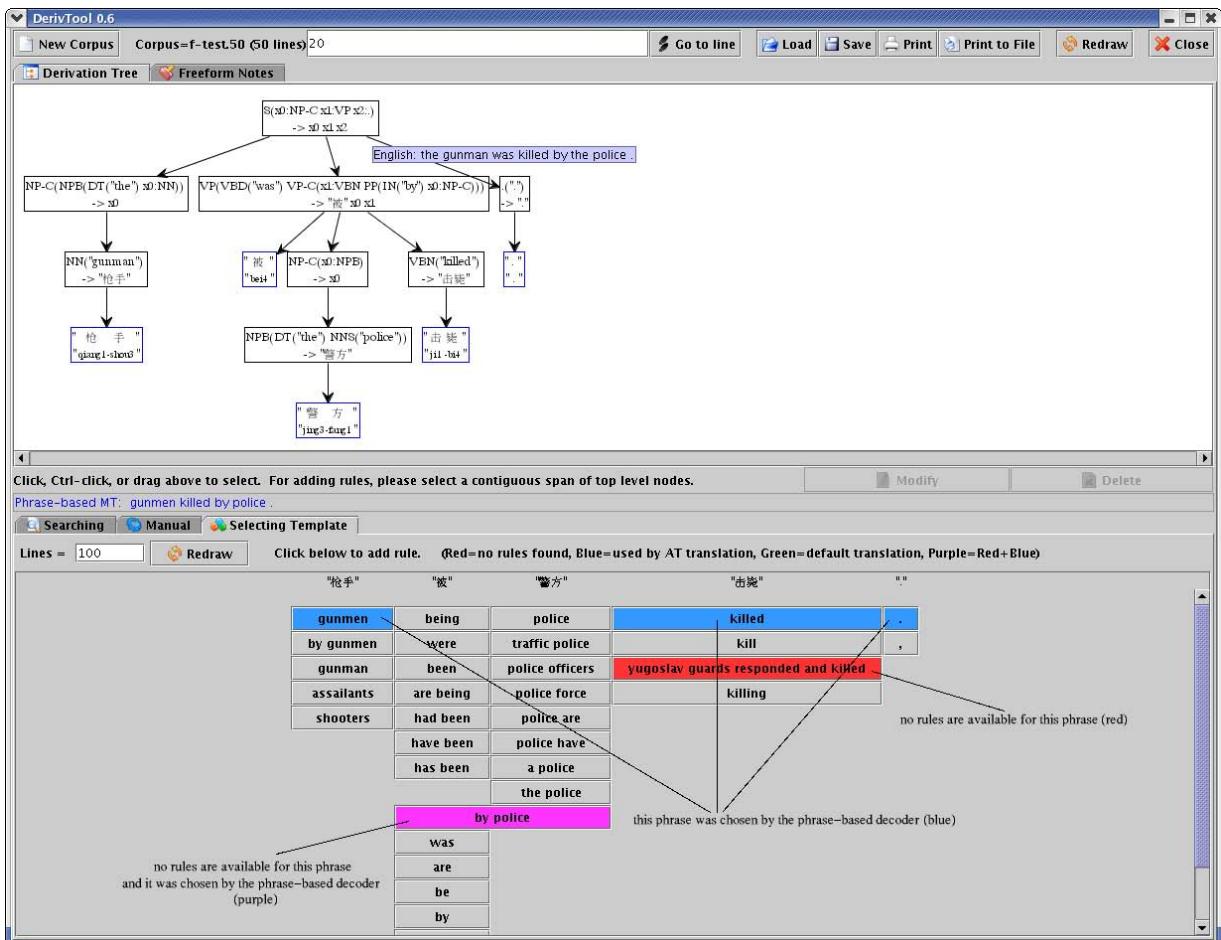


Figure 2.6: DerivTool visualizing the rule-based derivation of a machine translation for a sentence.

Linear B [29], shown in *Figure 2.7*, shows machine translations available for various spans of a sentence. It places the translations as spans underneath, in sections corresponding to the parts of the sentence that they translate. This visualization is perhaps the most intuitive to non-

experts of the ones discussed here, though it nevertheless still has the issue of requiring high levels of screen space. Unfortunately, while the paper proposes user studies for this interface, it does not actually report results.

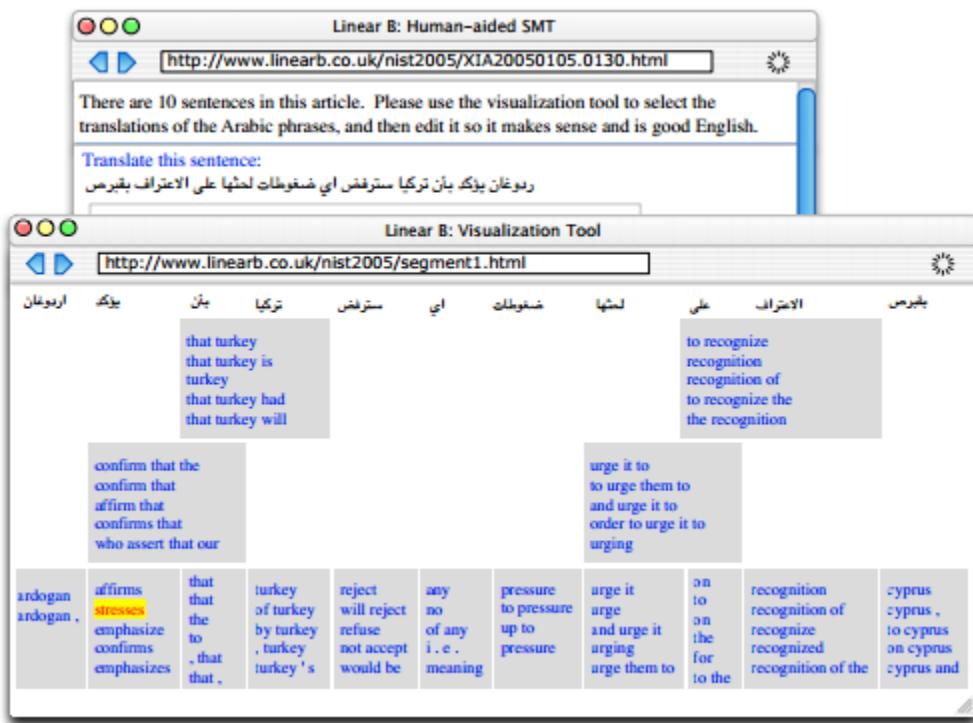


Figure 2.7: Linear B visualizing a sentence.

2.5 Webpage Reading Assistance

A number of tools are available to help users comprehend or learn from foreign-language webpages. One is Froggy GX [32], shown in *Figure 2.8*, which helps non-native English readers skim English webpages. The target user level for Froggy GX is much higher than the ones we are aiming for; these users are assumed to already know English grammar, and simply need the tool to improve their skimming speed and compensate for occasional vocabulary deficits. Indeed, the participants in the user studies for Froggy GX had over a decade of experience learning English on average.

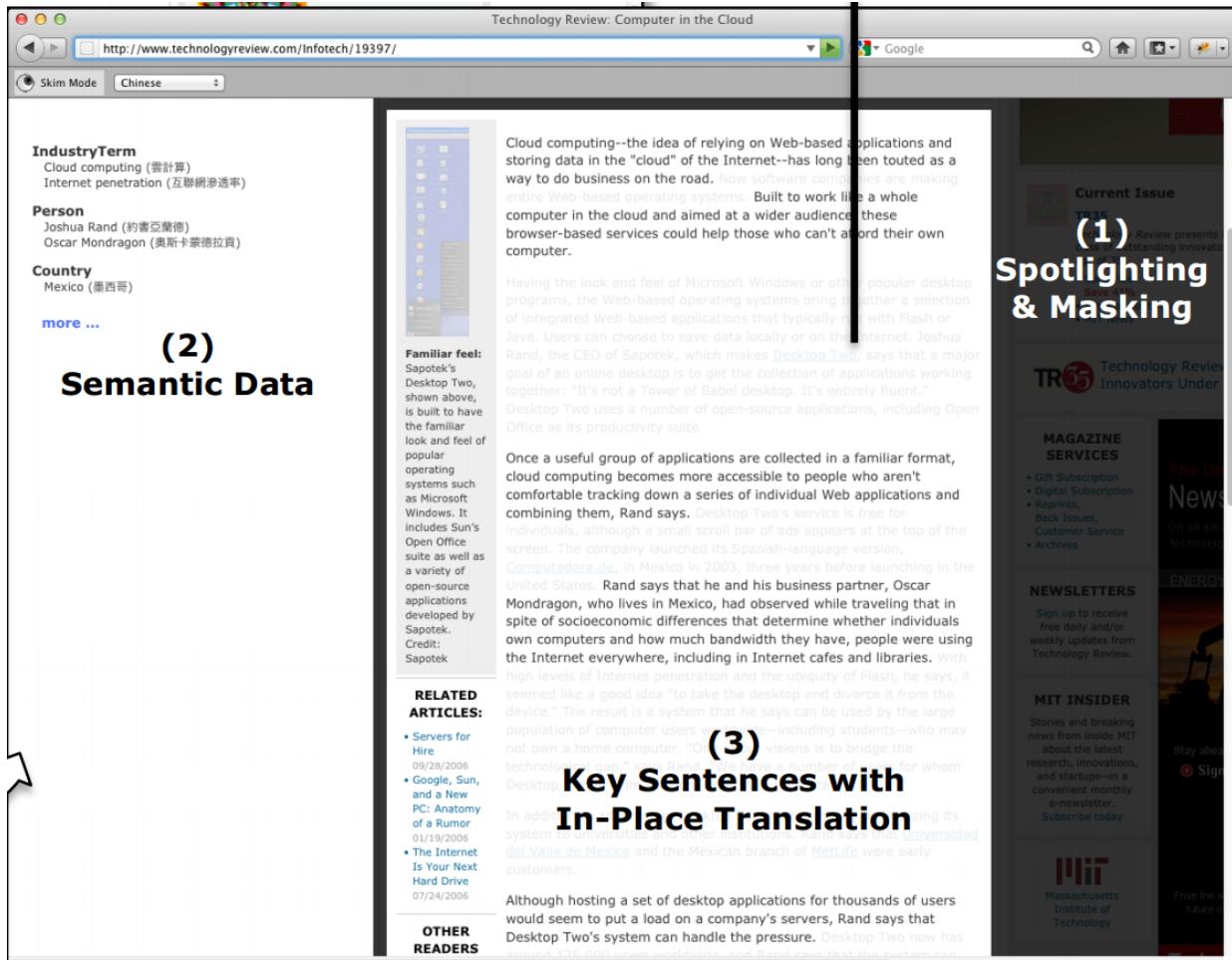


Figure 2.8: Froggy GX presents non-native English readers with translations and vocabulary to help them read.

Rikaikun [31], shown in *Figure 2.9*, is a Chrome extension that gives Japanese learners definitions for words as they hover over them in webpages. It is therefore particularly useful for learners who understand Japanese grammar, but encounter vocabulary or characters that they do not understand. That said, the system does not have any interface features geared towards visualizing or learning grammar.



Figure 2.9: Screenshot of the Rikaikun Chrome extension showing a word-level gloss on the Mainichi page.

This approach of providing word-level translations is quite prevalent among language-learning services. One such service is Duolingo [33], shown in *Figure 2.10*, where learners are often asked to translate sentences, and can get individual word-level translations by hovering over the word.

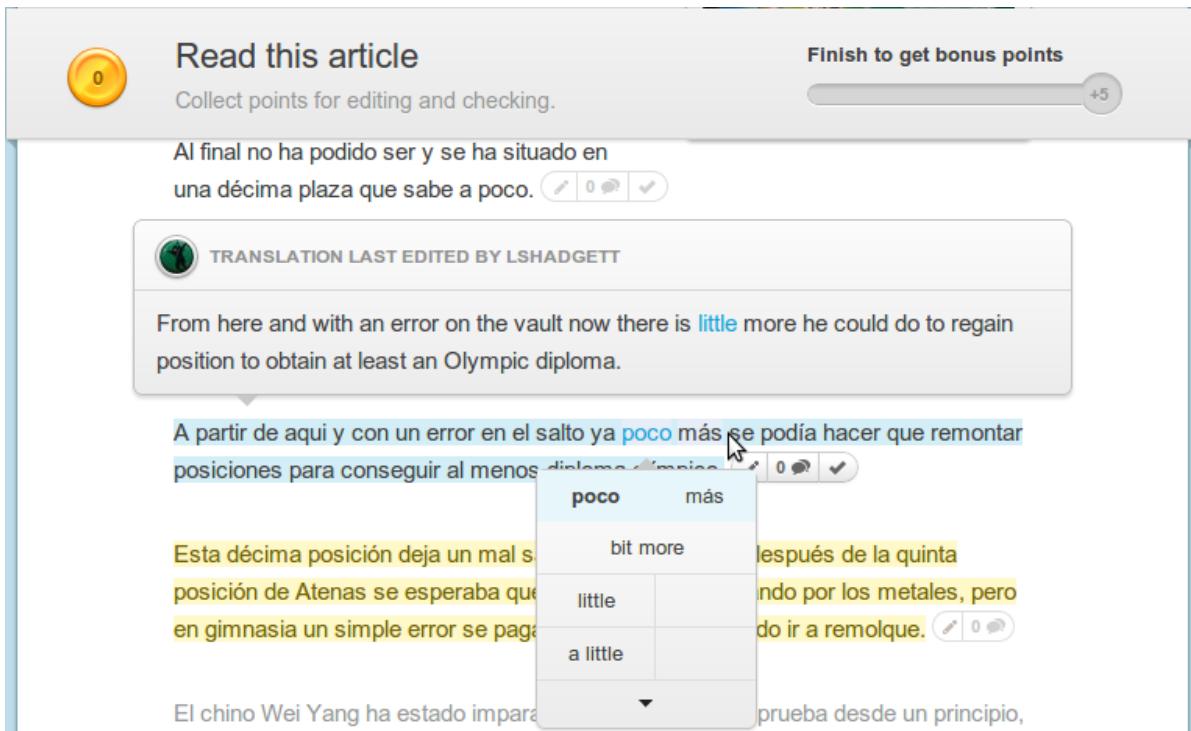


Figure 2.10: A word-level gloss being shown during sentence translation on Duolingo.

A common feature shared by all these approaches that rely on learners making grammatical inferences from word-level translations is that the sentences that are used for teaching must be relatively simple. Otherwise, the learner will not be able to comprehend the sentence and how the words relate to the overall meaning. Our grammar visualization addresses this issue by explicitly breaking down complicated sentences into short, simple phrases, on which the grammatical inference problem of relating words to phrases is less difficult for learners.

Chapter 3: Smart Subtitles

3.1 Interface

We developed a video viewing tool, **Smart Subtitles**, which displays subtitles to language learners to enhance the learning experience. It does so by providing support for dialog-level navigation operations, as well as vocabulary-learning features, which are shown in *Figure 3.1*. Smart Subtitles can be automatically generated for any video, provided that a caption is available.

3.1.1 Navigation Features

The navigation features of our interface were developed based on foreign language learners' video viewing patterns. In various informal interviews conducted with fellow language learners who enjoyed watching subtitled foreign-language videos for learning purposes, they reported that they often reverse-seeked to the beginning of the current line of dialog to review the portion that had just been said. Therefore, we aimed to make this process as seamless as possible. In our interface, clicking on a section of the dialog will seek the video to the start of that dialog, as shown in *Figure 3.2*.



Figure 3.1: The interface for Smart Subtitles. As can be seen from the bubbles, the interface provides features for both dialog navigation as well as vocabulary learning.



yuán lái zài zuò mèng

原来在 做梦

originally original; former; formerly; at first; so, actually, as it turns out

xiǎo jié

晓洁 翻译

kuàiqǐ chuáng bā diǎn sì shí fēn le lā

快 起床 八点四十分了啦

click to
seek



dì yī jí

第一集

yuán lái zài zuò mèng

原来在 做梦

翻译

originally original; former; formerly; at first; so, actually, as it turns out

xiǎo jié

晓洁

Figure 3.2: Our dialog navigation interface allows the user to click on any line of dialog to seek the video to that point. This helps learners re-listen to and review lines of dialog.

Another activity that some language learners reported doing was attempting to locate the line of dialog where a particular word or phrase had been said. Therefore, we enable easy seeking through the video based on dialog. The transcript is prominently shown, and can be navigated by pressing the up/down keys, or scrolling. It is also possible to search the video transcript for occurrences of particular words, as shown in *Figure 3.3*.



Figure 3.3: Smart Subtitles allows the user to scroll through a full transcript of the video to navigate the video. Additionally, the transcript can be searched for particular words or phrases, to locate their occurrences in the video.

3.1.2 Vocabulary Learning Features

The vocabulary learning features of our interface are aimed towards the use case where the viewer encounters an unknown word, and would like to look up its definition. The interface allows a user to hover over any word, and it will show the definition.

In addition, for languages such as Japanese and Chinese, which have non-phonetic writing systems, the interface also shows the phonetic representations for learners. For Chinese, it

shows pinyin, the standard romanization system for Chinese, as shown in *Figure 3.4*. For Japanese, it shows hiragana, the Japanese phonetic writing system, as shown in *Figure 3.6*.



Figure 3.4: Smart Subtitles showing pinyin for words.

For Chinese, when displaying pinyin, the pinyin is colorized according to the tone, to help make it more visually salient and memorable. The color scheme for tone colorization is based on the one used in the Chinese through Tone and Color series [8]:

Each of the four tones in Mandarin Chinese has been assigned a color:

1ST TONE RED: HIGH AND LEVEL
2ND TONE ORANGE: HIGH AND RISING
3RD TONE GREEN: LOW AND DIPPING
4TH TONE BLUE: HIGH AND FALLING

The neutral, unstressed tone is left black. Each entry contains important information about the character's tone and meaning. The page is printed in the character's assigned color, and a graphical representation of the tone's contour is in the upper right corner.

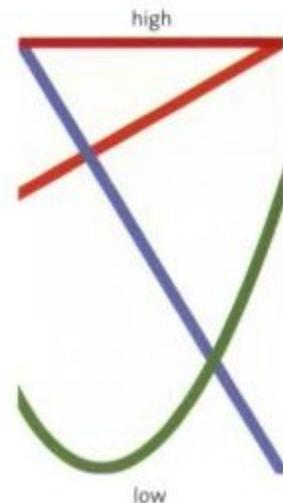


Figure 3.5: Color scheme from Chinese through Tone and Color [8]



いわや すみれ
(岩谷スミレ) そうですね→

さいごに な ちゅうがく じ
最後に泣いたのは 中学 の時→ 翻译

(adv) last; lastly; in conclusion; finally; JWN

し じ か
ベットが死んだ時です 風邪をこじらせてあっさり→

Figure 3.6: Smart Subtitles showing hiragana, the Japanese phonetic writing system, for words.

In this case, the さいごに displayed above the 最後に indicates that this word should be pronounced “saigoni”. The software additionally has an option to display in romanized form, but we display hiragana by default as it is the phonetic form that is generally presented, both in authentic materials, such as Japanese children’s books, in and foreign-language-learning materials

Finally, it may be the case that some sections of the dialog are idiomatic expressions that word-level translations will not be sufficient to comprehend. For example, just as English has idioms like “kick the bucket”, Chinese and Japanese have 成語 (set expressions). An example is 弱肉強食, literally, weak-food-strong-eat, meaning “survival of the fittest”. The full meaning of these set expressions cannot readily be determined from a word-level translation. Additionally, there may be certain grammatical patterns or grammatical particles whose meanings the learner has not yet learned, making such phrases difficult for learners to comprehend.

To address these cases where the word-level translations will not be enough for the learner to comprehend the dialog, we include an option for the learner to get a full translation for the

phrase, by pressing a button, as shown in *Figure 3.7*. The phrase-level translation can be obtained from a subtitle track in the viewer's native language if it was supplied, as in the case when we have a video with Chinese audio and we have both English and Chinese subtitles available. Alternatively, if we only have a transcript available, and not a subtitle in the viewer's native language, we rely on a machine translation service to obtain a phrase-level translation. Either Microsoft's or Google's translation service can be used. A later version of our interface, described in Chapter 4, describes an alternative way to address the unknown-grammar problem.



Figure 3.7: If the user clicks on the 翻译 (translate) button, a full translation for the current line of dialog will be displayed.

3.1.3 Excluded Features

We also developed various additional features, but opted to disable them based on negative user feedback during pilot testing.

One excluded feature was pausing the video automatically when the user hovered over a word. This was intended to give viewers additional time to read the definition. However, users remarked that it was distracting to have the video playback be interrupted, and seemed to significantly reduce their usage of the vocab-hover feature when this was enabled.

Another excluded feature was the integration of speech synthesis into the vocabulary hover feature. Namely, after a second of hovering over a given word while the video was paused, the pronunciation would be provided. This was intended to assist learners who might not yet be sufficiently familiar with the phonetic writing systems (pinyin or hiragana) of the foreign language, and could not discern the pronunciation from the audio, due to fast or incoherent speech. However, users remarked that the automatic audio playback feature was distracting and redundant given the pinyin display, and so we disabled it.

3.2 Implementation

Our system needs 2 sets of text-format subtitles: one in the original language of the video, and one in the viewer's language. We first describe systems we implemented to obtain these text-format subtitles from common video sources. Then, we will describe the process of generating the word-level annotations from the text-format subtitles.

3.2.1 Obtaining Subtitles for Videos

Subtitles for videos are available in a number of formats. They can be divided into the 3 categories: *text-based subtitles*, *bitmap-based subtitles*, and *hard subtitles*.

Text-based subtitles

Text-based subtitles, also often referred to as *soft subtitles*, consist of text, with an associated timestamp for when they should appear on the screen. Factors such as rich-text formatting and positioning of subtitles are controlled by keywords provided in headers or particular tags.

Examples of these formats include the SRT (SubRip) and WebVTT (Web Video Text Tracks) [13] formats. Text-based subtitles are the most convenient subtitle format from the perspective of computer-based processing, because the text content of them is directly machine-readable. Generally, text-based subtitles are available for download from online video-streaming sites, such as Amara or Viki.

In the SRT subtitle format, every block of 3 lines in the subtitle represents a line of dialog. The first line is the subtitle number, the second line is the start and end time of when this line will be displayed in the subtitle, and the third line is the text itself which will be displayed. An example is shown in *Figure 3.8*.

```
1
00:00:20,000 --> 00:00:24,400
Altocumulus clouds occur between six thousand

2
00:00:24,600 --> 00:00:27,800
```

and twenty thousand feet above ground level.

Figure 3.8: Example of an SRT format subtitle. This example displays the line “Altocumulus clouds occur between six thousand” from 20 thru 24 seconds in the video, and the line “and twenty thousand feet above ground level.” from 25 through 28 seconds in the video.

The WebVTT format is similar to SRT, except that it requires a line “WebVTT” as the first line in the file, does not require the initial subtitle numbers, and it permits a limited set of HTML tags within the subtitle for formatting purposes. These tags include formatting tags, such as the *<i>* tag to indicate that *<i>Laughs</i>* should be italicized, as well as tags to provide semantic information, such as the *<v>* tag to indicate the current voice / speaker. An example is shown in *Figure 3.9*.

```
WEBVTT

00:18.000 --> 00:20.000
<v Roger Bingham>And with me is Neil deGrasse Tyson

00:20.000 --> 00:22.000
<v Roger Bingham>Astrophysicist, Director of the Hayden Planetarium

00:22.000 --> 00:24.000
<v Roger Bingham>at the AMNH.

00:24.000 --> 00:26.000
<v Roger Bingham>Thank you for walking down here.

00:27.000 --> 00:30.000
<v Roger Bingham>And I want to do a follow-up on the last conversation we did.

00:30.000 --> 00:31.500 align:end size:50%
<v Roger Bingham>When we e-mailed—

00:30.500 --> 00:32.500 align:start size:50%
<v Neil deGrasse Tyson>Didn't we talk about enough in that conversation?

00:32.000 --> 00:35.500 align:end size:50%
<v Roger Bingham>No! No no no no; 'cos 'cos obviously 'cos

00:32.500 --> 00:33.500 align:start size:50%
<v Neil deGrasse Tyson><i>Laughs</i>

00:35.500 --> 00:38.000
<v Roger Bingham>You know I'm so excited my glasses are falling off here.
```

Figure 3.9: Example of a WebVTT format subtitle.

A final type of widely-used text-based subtitle are *Closed Captions*, which are often included in broadcast television. However, they are transmitted over-the-air and are generally converted into one of the above textual subtitle representations when captured by video capture cards.

Bitmap Formats

Bitmap formats do not include the machine-readable representation of the subtitle text. Rather, they are pre-rendered versions of the text which are overlayed onto the video when playing. Thus, they consist of a time-range associated with a bitmap image. Bitmap formats are the standard format used in DVDs, because it allows DVD producers to provide subtitles with arbitrary formatting in any language, while allowing the DVD player manufacturers to avoid providing fonts covering every possible character used in every possible writing system, in every possible font style that a DVD producer might require.

The particular bitmap-based subtitle format used in DVDs is called VobSub. It consists of an index file that indicates the timestamp at which each new subtitle should be displayed, as well as the offset in the subtitle-bitmap file at which the bitmap that should be displayed at that time is located. *Figure 3.11* shows an example index file.

```
# VobSub index file, v7 (do not modify this line!)
#
size: 720x480
palette: 000000, 000000, 000000, 000000, 000000, 000000, 000000, 000000,
000000, 000000, 000000, 000000, ccccccc, 000000, 000000
# ON: displays only forced subtitles, OFF: shows everything
forced subs: OFF

# Language index in use
langidx: 0
id: en, index: 0
timestamp: 00:00:00:022, filepos: 000000000
timestamp: 00:04:59:732, filepos: 000000800
timestamp: 00:05:01:029, filepos: 000001000
timestamp: 00:07:05:792, filepos: 000001800
timestamp: 00:07:05:814, filepos: 000002000
timestamp: 00:14:02:508, filepos: 000002800
```

Figure 3.11: An example of the VobSub-format index file which was extracted via a DVD ripper.

Of course, because the text of bitmap-based subtitles cannot be readily read from the index file, our system needs a pre-processing step in which it extracts the textual content from these subtitles. In practice, this is implemented as a separate program which converts the extracted index and bitmap files to an SRT-format subtitle. For each bitmap associated with each timestamp this program first applies OCR (Optical Character Recognition), a process which extracts text from images, to extract out the text. Then, it combines the text with the timestamps indicated in the index file to generate an SRT-format subtitle. Our implementation uses Microsoft OneNote [42] as the OCR engine, via the Office Automation API [43]. We chose OneNote due to its high OCR accuracy on Chinese-language subtitles. However, the free Tesseract [44] OCR engine can also be used instead.

Hard Subtitles

Yet another way that subtitles may be provided is via hard subtitles. Hard subtitles include the subtitle as part of the video stream. This approach has the advantage that it has no software requirements whatsoever to be used - any player that can play video, will also be able to display the hard subtitles (since they are part of the video). Of course, this approach has the disadvantage that they are non-removable. Additionally, hard subtitles are the most difficult to get machine-readable text out of, because the subtitle must first be isolated from the background video, before we can apply OCR to obtain the text. That said, hard subtitles are ubiquitous, particularly online - videos of Chinese-language dramas on video-streaming sites such as Youku are almost always accompanied by a hard-subtitled set of subtitles. Thus, in order to allow our system to be used on these types of videos as well, we have devised an algorithm which can nevertheless identify the subtitle from hard-subtitled videos and extract it out, which is described in the following section.

3.2.2 Extracting Subtitles from Hard-Subtitled Videos

Summary

Subtitle extraction from hard-subtitled videos is a challenging task. It aims to extract text from the video, which is in an unknown font, in an unknown color, at an unknown region of the screen. In this chapter, we will describe and evaluate an algorithm we have developed for extracting subtitles from Chinese-language hard-subtitled videos.

Extracting the subtitles from a hard-subtitled video consists of the following general steps:

- 1) Identify the frames in which a subtitle actually appears
- 2) For the sets of frames on which a subtitle is displayed, isolate out the regions with text
- 3) Apply OCR to get the text.

These general stages are not as clearly separated in practice as we have described. For example, if in the 1st stage we mistakenly select a frame in which a subtitle doesn't exist, the failure of the OCR process in the 3rd stage will allow us to correct this error.

Manual Process

There exist a few tools which assist with the hard-sub-extraction procedure, though they require human intervention to be used, and therefore cannot be directly compared to our fully-algorithmic approach. One such tool is SubRip (the same tool after which the SRT format is named).

The screenshots used in the following overview of the hard-subtitle extraction process via SubRip are from [40].

When using SubRip to extract subtitles out of hard-subbed videos, the user first specifies a box in which the subtitles are expected to appear, then scans through the video until he finds a frame where a subtitle appears, as shown in *Figure 3.12*.

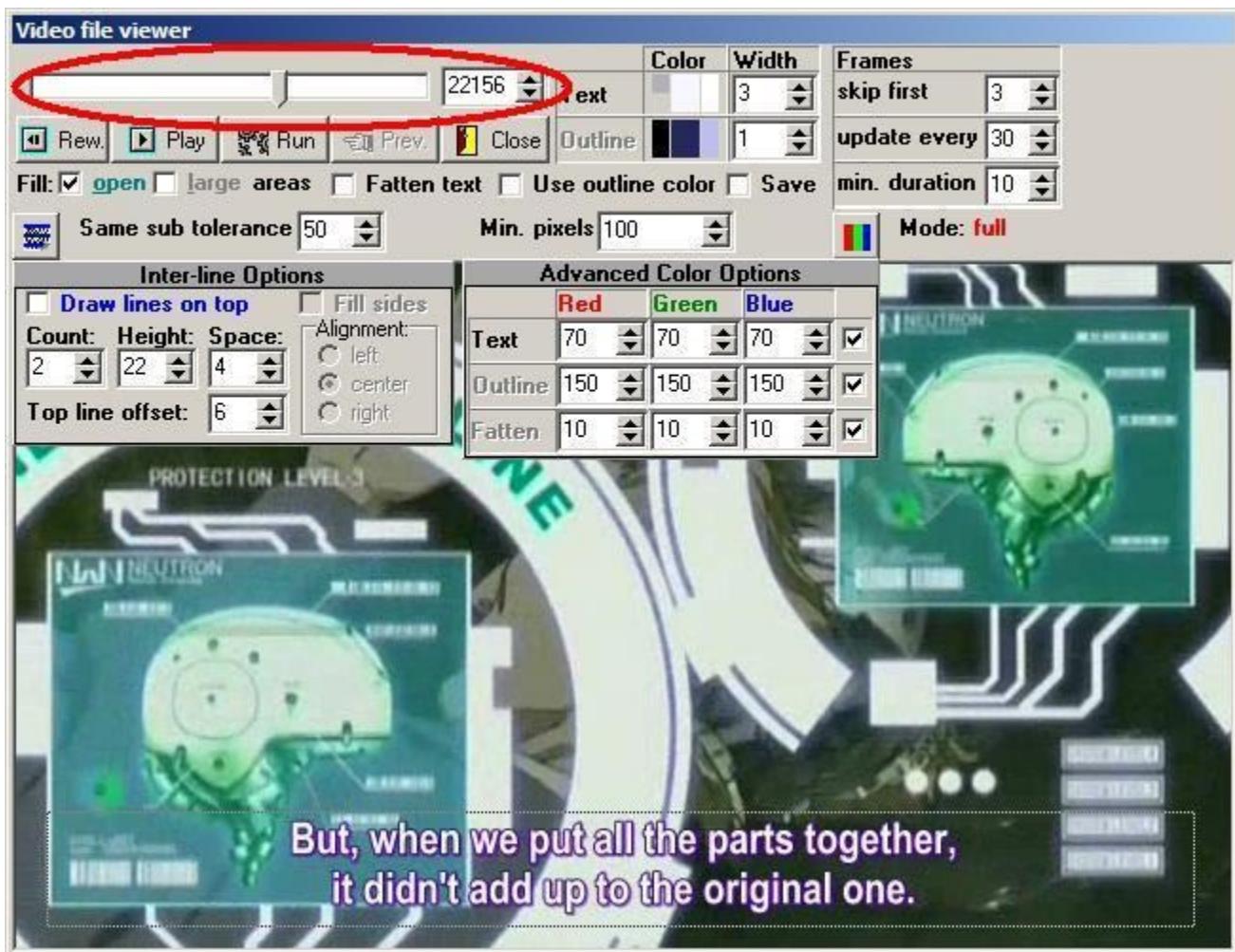


Figure 3.12: Using SubRip to seek to a region of the video that contains subtitles, and marking the region where the subtitle appears.

Next, the user specifies the color of the subtitle. This eliminates pixels which are not of that color or similar to it, as shown in *Figure 3.13*.

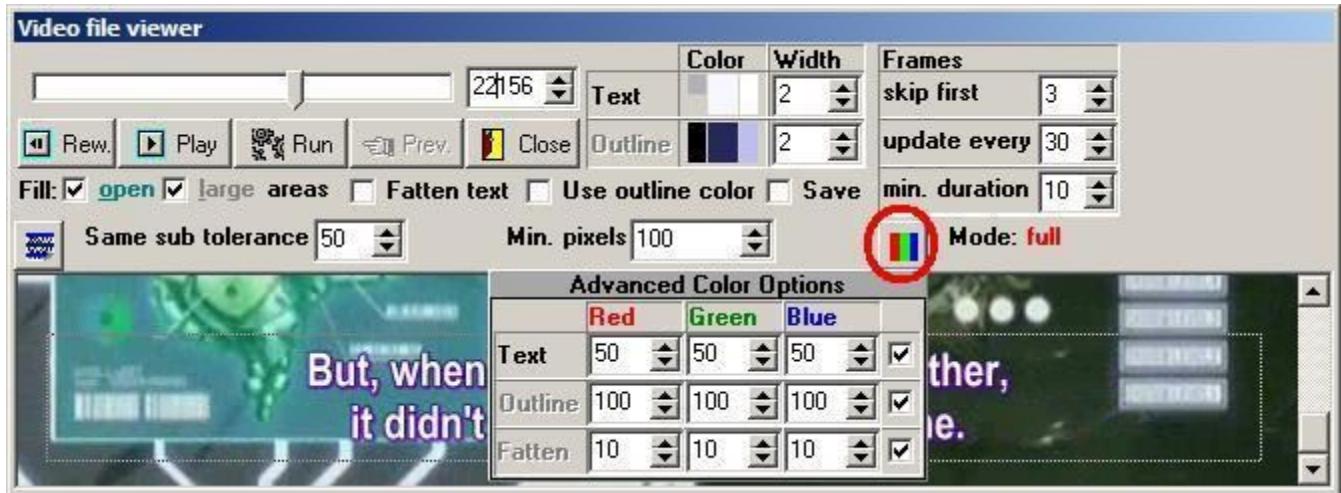


Figure 3.13: Using SubRip to specify the subtitle color.

Finally, the user specifies the boundaries of the particular lines in the subtitle. This eliminates extraneous pixels which happen to be of the same color, but are outside the subtitle boundary, as shown in *Figure 3.14*.



Figure 3.14: Using Subrip to specify the boundary of the subtitle.

After this process, we have isolated the subtitle text from the rest of the video in this frame. We now have an image on which we can apply OCR, as shown in *Figure 3.15*. SubRip's implementation uses a template-based matching system which is font-specific, and does not work on languages using non-Latin scripts, such as Chinese or Japanese. However, one could, in principle modify it to use a general-purpose OCR engine such as OneNote's.

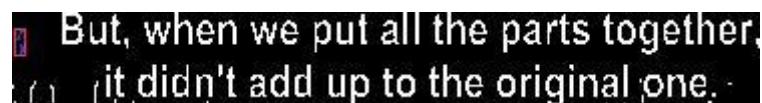


Figure 3.15: Image following SubRip's subtitle isolation process, on which OCR can then be applied.

Thus, as we can see, the process of doing subtitle extraction from hard-subbed videos using tools such as SubRip requires high amounts of manual intervention, and so we opted to develop a system which will automate many of these processes.

Our Implementation

Our implementation makes a few assumptions about subtitles which we have found generally hold true in the Chinese-language hard-subbed material we have observed:

- 1) Subtitles in the same video are of the same color, with some variance due to compression artifacts.
- 2) Subtitles in the same video appear on the same line.
- 3) Subtitles appear somewhere in the bottom vertical quarter of the screen.
- 4) Subtitles are horizontally centered. This approach will not work unmodified with subtitles that are left-aligned.
- 5) Subtitles do not move around and are not animated, and will remain static on-screen for a second or more to give the viewer time to read them.
- 6) Characters in the subtitle generally have the same height. This is a Chinese-specific assumption, as Chinese characters are generally of uniform size.
- 7) Characters in the subtitle have lots of corners. This is also a Chinese-specific assumption, owing to the graphical complexity of Chinese characters.

The steps that our implementation uses to extract the subtitles are:

- 1) Determine a small region of the video in which some part of the subtitle is most likely to be present
- 2) Determine the color of the subtitle
- 3) Determine the vertical span of the subtitle line
- 4) Eliminate connected components spanning outside the subtitle (disabled feature)
- 5) Determine the horizontal span of the subtitle line
- 6) Restrict to pixels which remain static across frames
- 7) Run OCR on the remaining frames to extract out the text content of the subtitles
- 8) Eliminate duplicates and false positives, and output the final subtitles



Figure 3.12: The result of applying a corner detector to the top image. As we can see, Chinese characters have many corners, which make the corner detector an effective means of detecting where the Chinese characters are. Extracting out the white-colored pixels that are near corners results in the bottom image, which the OCR engine is able to extract text out of.

Finding a region of the video where subtitles are likely present

The first step our tool makes is to determine a small region of the video where subtitles are definitely believed to be present. This will be used in later stages to determine the color of the subtitle, and the frames in which subtitles are present. The size of the region we aim to find is 1/32 of the vertical size of the video, and 1/4 of its horizontal size.

To find this region, the algorithm first vertically crops the video to the bottom quarter, making use of our assumption that the subtitles are somewhere towards the bottom. It also horizontally crops the video to the middle quarter, making use of our assumption that subtitles are horizontally centered. Then, it runs a corner detector across the frames of the video. The corner detection is done via the Harris corner detector [45], implemented in OpenCV [46]. As shown in *Figure 3.12*, the corner detector has high values at regions where Chinese characters are present, due to their graphically complex nature. Then, for each of the frames, the algorithm splits it vertically into 8 regions, and in each of these regions it computes the sum over the corner detection values. It then selects the region where this value is maximized across all frames. Thus, at the end of this process, we have determined the 1/32 vertical region in which we believe a subtitle is most likely to be present.

Determining the color of the Subtitle

Next, we determine what the color of the subtitle will be. We do this by considering the 1/32 by 1/4 region in which the subtitle is believed to occur. Within this region, over all frames, we consider all pixels near the regions where the corner detector output high values (which we believe to be the pixels at which the Chinese characters are), and tally the colors of these pixels. We ignore black or near-black color values, because the interior of subtitles are almost always in a light color, and we wished to avoid getting the color of the subtitle outline or background instead of the subtitle itself. Colors are rounded to the nearest 12-bit value, meaning that there are 16 possible values for each of the red, green, and blue color channels. We round colors to these values, as opposed to considering them in their original 24-bit colorspace, because compression artifacts and slight gradient effects within the subtitle will result in the color value being slightly different between pixels.

Determining the vertical span of the subtitle line

One might assume that, once we have determined the subtitle color, we can simply extract out the pixels that roughly match that color to isolate out the subtitle. However, this will result in high levels of noise, particularly if parts of the background match the subtitle color. For example, for the image in *Figure 3.12*, extracting out the pixels which roughly match the subtitle color results in an image with extraneous artifacts:



Figure 3.13: Simply extracting out the pixels which match the subtitle color results in an image with extraneous artifacts.

For a more extreme set of examples, we can consider the image in *Figure 3.14*, where the results of simply extracting out pixels which roughly match the subtitle color will clearly not be OCR-able:



Figure 3.14: Simply extracting out the pixels which match the subtitle color results in an image with extraneous artifacts.

If we also apply our near-corners criterion in addition to the color restriction, we can eliminate some of the noise, but not all of it, as shown in *Figure 3.15*.



Figure 3.15: Restricting pixels to those that both match the subtitle color and are near corners reduces extraneous artifacts, but does not always eliminate them.

Top: original image

Middle: corner detection results

Bottom: pixels which match the subtitle color and are near corners

To reduce the amount of noise further, we therefore wish to make use of additional cues based on knowing that subtitles have uniform height, which the subtitle text does not extend above or below. To do so, we must first determine what the height of the subtitle is.

Let us define *hot pixels*, or pixels where we believe a subtitle is likely present, as pixels that match the subtitle color and are near corners. Our algorithm works by first restricting attention to those frames in the video where a subtitle is likely to be present. These are the 1/2 of frames that contain the most number of hot pixels, as we assume that subtitles are present in less than half of the frames. Next, we find the vertical section of the subtitle that maximizes the ratio of total number of hot pixels encompassed to the height of the area. We add a small bonus for each additional percentage of the height introduced, to ensure that the selected region will encompass as large a portion of the subtitle as possible. At this point, we have identified the vertical region in which the subtitle is present, so we eliminate the hot pixels which fall outside this vertical span.

Eliminate connected components which extend outside the subtitle boundary (disabled feature)

Another observation we can make is that individual letters in the subtitle also often have some form of shadow or border that separates them from the background. Therefore, if we are considering a blob of subtitle-colored pixels, if it extends outside the vertical region in which the subtitle is supposed to be in, then it probably is not a letter but rather part of the background.

These “blobs” of continuous-color regions are generally referred to as *connected components*, shown in *Figure 3.16*. We can efficiently find connected components using connected component labelling algorithms [47].



Figure 3.16: An illustration of connected components. The bottom image displays each connected component present in the top image in a unique color.

After finding the connected components, we consider the topmost and bottommost portions of each connected component, and if it falls outside the subtitle boundary, then we assume that the connected component is not a letter, and therefore eliminate all pixels that are present in it.

That said, this approach crucially depends on there existing a distinct outline of a different color around the subtitle. We found that in certain videos, this is not present, resulting in this stage eliminating parts of the letter, as displayed in *Figure 3.17*:



Figure 3.17: The '尔' part of 你 was eliminated as part of the connected component-based elimination process, which will cause an OCR error later on. (尔 is a different character from 你)

Therefore, because of the errors it introduces when subtitles lack a border, and the fact that this noise can be eliminated in later stages as well, we opted to disable this connected-component based feature in our final algorithm.

Determine the Horizontal Span of the Subtitle Line

At this point, we have identified the vertical strip in which the subtitle is located, as well as the subtitle color, and have eliminated pixels falling outside the subtitle's vertical region. Now, we can eliminate some additional extraneous pixels that might be at the sides by identifying where the subtitle is horizontally located, as shown in *Figure 3.18*. To do this, we take advantage of our assumption that the subtitles are horizontally centered, and apply the same methodology as we did for identifying the vertical subtitle span. We start at the middle, expanding equally outwards, and find the horizontal section of the subtitle that maximizes the ratio of total number

of hot pixels encompassed to the width of the area. As before, we give a small bonus for each additional percentage of width introduced, to ensure that the selected region will encompass as large a portion of the subtitle as possible. At this point, we can eliminate the hot pixels that fall outside this horizontal region.



Figure 3.18: Determining the horizontal span of the subtitle further eliminates some extraneous pixels.

Top: original frame

Middle: hot pixels that fall within the detected vertical subtitle boundary

Bottom: hot pixels that fall within the vertical and horizontal subtitle boundaries

Restrict to Pixels which remain static across frames

The noise-reduction algorithms we have described thus far mostly operate within the same frame, with the exception of determining the subtitle color and vertical position, which is computed over multiple frames. However, we can additionally take advantage of the fact that a subtitle will remain on-screen for at least a second so that a reader will have time to read it, over which multiple frames will occur. At 30 frames per second, this would be over 30 frames. Thus, since the background in videos may change, but the pixels for the subtitle itself will remain constant, this is an additional cue that we can use to further eliminate extraneous pixels, as shown in *Figure 3.19*.



Figure 3.19: An example of why it is useful to consider multiple frames: these 2 frames both have the same text, but the second frame will have less extraneous noise after applying the noise-removal steps we described above. The pixels shared by both frames are the pixels that make up the subtitle text.

First, we wish to group together frames which we believe have the same subtitle text showing in them. To do so, we compute between each pair of frames a similarity metric: the number of hot pixels which remain preserved between the frames, divided by the maximum of the number of hot pixels in either of the two frames. This is therefore a metric between 0 and 1; if the value is above 0.5, we consider the two frames to have the same subtitle text. By computing this metric on each consecutive frame, we can determine where the subtitle transitions occur, and thereby group together the subtitles which are common. Note that we did not have any explicit step in determining whether a frame actually contains subtitles in the first place. We will act as though they do for the time being, and eliminate those groups that don't contain a subtitle in later stages.

Now that we have grouped together frames containing the same subtitle text, we now wish to determine a representative image from these which will yield good results when OCR-ed. Since our OCR process is slow, it is unfortunately not feasible to simply try to OCR each frame and eliminate the bad frames based on the OCR output.

One might attempt to simply take the intersection of all of them, leaving only the pixels that are shared in common. However, as we showed earlier, some of our heuristics eliminate pixels which actually are part of the subtitles in certain conditions, so taking an intersection will lead to suboptimal results, as shown in *Figure 3.20*.



Figure 3.20: This figure illustrates why we should not simply take an intersection over all frames to determine the final set of pixels which we will use for OCR. On the first image, applying our subtitle-isolation heuristics in isolation results in a good set of pixels that can be OCR-ed to obtain the subtitle. However, our subtitle-isolation heuristics perform poorly on a later frame (third image), so simply taking the pixels which are in common across all frames will result in a set of pixels which are unusable for OCR.

The approach we opted for is to take the average of the pixel values. This is done after converting to greyscale and inverting the color, since the OneNote OCR engine appear to function best on black-on-white, as it is the usual color scheme of scanned, printed documents. However, this results in the various noise artifacts that appeared in only a handful of images still appearing faintly in the final image, which we found to cause issues at the OCR stage. Therefore, as a final post-processing step before passing the image to the OCR engine, we eliminate any hot pixels which only appeared in less than 1/4 of all of the frames, convert the color scheme to black-on-white, and increase the contrast ratio, as shown in *Figure 3.21*.



Figure 3.21: The end result of applying our algorithm on a set of frames. The final output is OCR-able.

Run OCR on the remaining frames to extract out the text content of the subtitles

At this point, we have a characteristic image containing the text in each of our groups of frames. Now, we run OCR (Optical Character Recognition) on each of them to extract out the text content from the images. A number of OCR engines are available, including Tesseract [44], an open-source engine. However, we choose to use OneNote [42] because it is the most affordable engine which has good recognition results on Chinese and Japanese subtitle text.

Eliminate Duplicates and False Positives, and output Final Subtitles

Now, we have the OCR output for each of the frame groups, which we need to do some post-processing on before we can output it into an SRT-format subtitle. First, we take the frame groups for which the OCR system was unable to detect text, and eliminate them (we assume that these were sections of the video where no subtitle was shown). Then, we take adjacent frame groups, and compare whether the OCR output was identical; if so, we merge them. Finally, we output the text and timestamps as part of the SRT file.

Evaluation

To evaluate this system, we tested the hard-sub extraction process on a 4 hard-subbed Chinese videos. As we did not have access to reference subtitles for these videos, and generating them manually would be prohibitively expensive, our observations will be mostly qualitative. Results

depend heavily on the video resolution and font. The system generally tended to work better on higher-resolution videos, and videos whose fonts tended to be thicker.

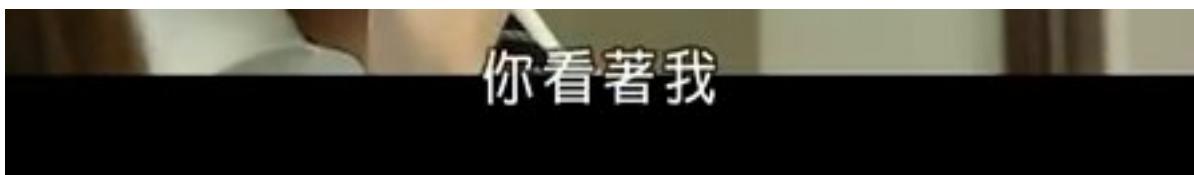
On one of the videos, the subtitle-color detection portion of the algorithm detected the wrong subtitle color, selecting the color of the grey subtitle outline rather than the white interior. Because the incorrect subtitle color was detected, the resulting subtitles were entirely incorrect. For videos such as these, the option to manually override the algorithm for certain critical sections, such as the color selection, would be helpful.

On the remaining videos, the detection of the subtitle color and region was correct. Most errors resulted from OneNote reporting that no text was detected on images that had extraneous noise, such as the one in *Figure 3.22*. This issue could potentially be resolved if we were able to customize the OCR engine to aim for a higher recall score, and attempt to recognize noisy characters. This issue was the most common of all, and occurred on roughly 10-30% of the dialogs, depending heavily on the video. Videos with white subtitles and white backgrounds, such as snow and actors wearing white clothing, tended to have the most noise.

重要的事要交付于你

Figure 3.22: OneNote's OCR engine will often not produce any output for images that have extraneous noise

Our system would also sometimes eliminate thin parts of characters that were in straight sections, far from corners. This likewise leads to OneNote not producing any OCR output. This issue occurred primarily in low-resolution videos. Addressing this issue may require us to adjust parameters to ensure the algorithm features related to corner proximity are independent of subtitle thickness and video resolution.



·你看著我

Figure 3.23: In the second character, the middle-bottom strokes were apparently too far from corners, and were eliminated, leading to an OCR error

The character-level error rate was quite low in the dialogs we looked at. Generally, character-level errors were restricted to a handful of characters which were repeatedly misrecognized. For example, in one video, 天, a common character meaning sky or day, was repeatedly misrecognized as 夂. This is, of course, an issue of the OCR system, which we cannot address.

Our assumption of a uniformly colored, horizontally centered subtitle location did not always hold outside the main video dialogs. For example, the subtitle used in opening themes was often in a different location and color from that used in the main video, as shown in *Figure 3.24*. The opening themes also sometimes had heavily stylized or animated characters that did not lend themselves to being OCR-ed, as shown in *Figure 3.25*. These same issues tended to also occur during musical asides, where the subtitles displayed lyrics of background music, as shown in *Figure 3.26*.



Figure 3.24: This opening theme violates our assumption of uniform subtitle color and size.



Figure 3.25: In this opening theme, the subtitle is animated karaoke-style, and is not static.



Figure 3.26: In this musical aside, the subtitle is not centered.

That said, since the main dialogs provide language learners with the most material in the context of learning from video are the main dialogs, extracting lyrics from opening themes and musical asides is a lower-priority goal. Although our subtitle extraction system has shortcomings, we believe it can be a useful tool for obtaining text-format subtitles for usage with the Smart Subtitles system.

3.2.3 Getting Words, Definitions, and Romanizations

Listing Words

Subtitles provide us with the text of each line of dialog. For many languages, such as English, going from these sentences to the list of words it includes is quite simple, since words are delimited by spaces and punctuation. However, there are a few corner cases such as words like “Mr.” that must be handled separately. This process of getting the words in a sentence is known as *tokenization*, and it can be done by a *tokenizer*, such as the one included in the Natural Language Toolkit (NLTK) [15]. Thus, using a tokenizer, I can input a sentence such as:

Mr. Brown ate at Joe’s, but didn’t like it.

And this will output a list of words and punctuation which occur in the sentence, which are referred to as *tokens*:

[‘Mr.’, ‘Brown’, ‘ate’, ‘at’, ‘Joe’s’, ‘,’ ‘but’, ‘didn’t’, ‘like’, ‘it’]

A particular issue which occurs with Chinese and Japanese is that the boundaries between words are not indicated in writing (whereas in languages such as English, spaces indicate the boundary between words). Thus, since our vocabulary-hover feature requires us to know what the words themselves are, we need a means to determine what the boundaries between the words are. For this purpose, we use statistical word segmenters for Chinese and Japanese - in particular, we use the Stanford Word Segmente [14] for Chinese, and JUMAN [17] for Japanese.

Getting Definitions and Romanizations

For languages such as Chinese that lack conjugation, the process of obtaining definitions and romanizations for words is simple: we simply look them up in a bilingual dictionary. The dictionary we use for Chinese is CC-CEDICT [19]. This dictionary provides both a list of definitions, as well as the pinyin for each word.

The process of listing definitions is more difficult for languages that have extensive conjugation, such as Japanese. In particular, bilingual dictionaries (such as WWWJDIC [20], which is the dictionary we use for Japanese) will only include information about the infinitive (unconjugated) forms of verbs and adjectives. However, the words which result from segmentation will be fully conjugated, as opposed to being in the infinitive form. For example, the Japanese word meaning “ate” is 食べた [tabeta], though this word does not appear in the dictionary; only the infinitive form “eat” 食べる [taberu] is present. In order to provide a definition, we need to perform *stemming*, which is the process of deriving the infinitive form a conjugated word.

Rather than implementing our own stemming algorithm for Japanese, we adapted the one that is implemented in the Rikaikun Chrome extension [31].

For other languages, such as German, instead of implementing additional stemming algorithms for each language, we instead observed that Wiktionary [41] for these languages tends to already list the conjugated forms of words with a reference back to the original, and therefore we simply generated dictionaries and stemming tables by scraping this information from Wiktionary.

3.2.4 Translation-sense Disambiguation

去 [qu4]

Definitions:

1. to go
2. to go to (a place)
3. to cause to go or send (sb)
4. to remove
5. to get rid of
6. (when used either before or after a verb) to go in order to do sth
7. to be apart from in space or time
8. (after a verb of motion indicates movement away from the speaker)
9. (used after certain verbs to indicate detachment or separation)
10. (of a time or an event etc) just passed or elapsed/

Figure 3.27: The word 去 has many possible translations to English. Which of them is the best translation? This is the problem of Translation-sense Disambiguation.

jié guǒ què diē tíng le sān tiān

结果却跌停了三天

翻译

result to bear fruit; CL:個|个[ge4]; outcome; conclusion; in the end; as a result; to kill; to dispatch

Figure 3.28: Translation-sense disambiguation can be used to determine which of the dictionary definitions should be placed at front (the remaining definitions follow and are greyed-out).

Smart Subtitles show word-level translations for words, however the correct translation for an individual word, also known as its *translation sense*, depends on the context in which it appears. Determining the correct translation for a given word in context is the problem of *Translation-Sense Disambiguation*. As there were no software packages readily available that perform this task for Chinese, we decided to implement our own.

Below, we will describe an overview of the translation-sense disambiguation algorithm we developed and its evaluation. Refer to [39] for additional details on its implementation, training, evaluation, and related work on translation-sense disambiguation.

Algorithm Description

For each Chinese word with multiple possible translations to English, we trained an SVM (support vector machine) classifier to determine the correct translation sense. The classifier takes as input a sentence and a word that appears in it, and outputs the index of the correct dictionary definition for the particular context.

The classifiers were trained on training data that labelled words in a Chinese sentence with a reference English translation from the CC-EDICT dictionary [23]. This data was generated from a bilingual corpus [22] via a procedure described in [39].

The features extracted from the sentence for use in translation sense classification are numeric features based off co-occurrences of words, their distances in the sentence, and part of speech tags. They are described in further detail below.

Feature: Co-occurrence of Word and Feature Word

This feature considers the set of Chinese words that co-occur with the word in the training set sentences. Co-occurring words are words that occurred in any sentence that contained the word we're doing translation-sense disambiguation for. The feature vector is only those sets of words that co-occurred in the training sentences, as opposed to the complete vocabulary, to limit the dimensionality. To extract this feature, we look at whether any of the words in the surrounding sentence includes the feature words. This feature is boolean: 1 if the word is present, 0 otherwise. It does not consider the number of times the word appears.

Feature: Distance between Word and Feature Word

The word co-occurrence feature does not take into account how far the feature word was from the word being classified. Presumably, feature words that tend to occur closer to the word being classified are stronger signals. This feature takes the distance into account by considering the number of words between the word and feature word, and dividing by the number of words in the sentence. If the word does not co-occur, it gets a score of 1, as though it had been the maximal distance away from the word.

Feature: Fraction of Words in Sentence that is the Feature Word

The word co-occurrence feature does not take into account the number of occurrences of the feature word, or the length of the sentence. Presumably, the more occurrences of a feature word relative to the sentence length, the stronger a signal it is. This takes occurrence count into account by having the value for the feature not be a boolean, but rather the fraction of words in the sentence that was that feature word, which is a real value between 0 and 1.

Feature: Co-occurrence with Highly Co-occurring Words

We found upon implementing the features above that the feature vectors tended to be very sparse in the test case examples, simply because most of the words that co-occurred in the training set sentences didn't directly also occur in the test set sentences. However, words that are used in the similar contexts to the feature words may have occurred in the test set sentences. We defined a notion of word usage similarity by the number of co-occurrences in the training set:

$$\text{word_similarity}(\text{word1}, \text{word2}) = \frac{\# \text{of sentences in which words 1 and 2 both occurred}}{\# \text{of sentences in which either word 1 or 2 occurred}}$$

Clearly, `word_simularity` for any pair of words ranges between 0 and 1; if `word1` and `word2` are equal or always used in the same sentences, `word_simularity` is 1; if they are never used in the same sentence, `word_simularity` is 0.

To incorporate this notion of word similarity into our features, we have the value of the word similarity be the sum over the word similarities in the sentence.

$$\text{score}(\text{feature}) = \sum_{\{w \text{ in sentence}\}} \text{word_similarity}(w, \text{feature}) / \text{len(sentence)}$$

Feature: Part-of-Speech tags

We were also able to derive part-of-speech tags from the Stanford part-of-speech tagger for Chinese [5], which we used as features. Specifically, our features were the part-of-speech tag of the word that was being classified, as well as the part-of-speech tag of the surrounding words

Algorithm Effectivness

We evaluated our algorithm by comparing its ability to determine the correct translation sense, against a baseline heuristic of choosing the most frequently used translation sense. The test data on which we evaluated were 4285 Chinese words that occurred in sentences, which were labelled with a reference translation sense from the CC-EDICT dictionary via a procedure described in [39].

Choosing the most frequent translation sense for each word resulted in the reference translation senses for 45% of the words. Our algorithm performed better than this baseline, correctly determining the translation senses for 55% of words.

While our algorithm arrived at the reference translation sense for only 55% of the translation sense correctly, many of the errors were simply due to unnecessary definition distinctions made by the bilingual dictionary. For example, CC-EDICT listed both “TV” and “television” as separate translations for 电视, even though the definitions can be interchangeably used. We observed that our implementation had more significant gains in translation-sense classification

performance relative to our baseline heuristic on words where there was a more meaningful distinction between definitions [39]. Therefore, we believe that our translation-sense disambiguation system is suitable for use in the Smart Subtitles system.

3.3 User Study

Our user evaluations for Smart Subtitles was a within-subjects user study that compared vocabulary learning with this system, to the amount of vocabulary learning when using parallel English-Chinese subtitles. Specifically, we wished to compare the effectiveness of our system in teaching vocabulary to learners, compared to dual subtitles (which are believed to be the current optimal means of vocabulary acquisition during video viewing [3]).

3.3.1 Materials

The video we showed was the first 5 minutes, and the next 5 minutes of the drama 我是老师 (I am a Teacher). This particular video was chosen because the vocabulary usage, grammar, and pronunciations were standard, modern spoken Chinese (as opposed to historical dramas, which are filled with archaic vocabulary and expressions from literary Chinese). Additionally, the content of these video clips - various conversations in a classroom and household setting - was everyday, ordinary settings, so while there was still much unfamiliar vocabulary in both clips, cultural unfamiliarity with the video content would not be a barrier to comprehension. The Chinese and English subtitles were extracted from a DVD and OCR-ed to produce SRT-format subtitles.

3.3.2 Participants

Our study participants were 8 students from the Chinese 3 class at MIT. This was the third semester of Chinese at MIT, and our study was conducted at the end of the semester, so participants had approximately 1.5 years of Chinese learning experience. 4 of our participants were male, and 4 were female.

3.3.3 Research Questions

The questions our study sought to answer were:

- 1 Will users learn more vocabulary using Smart Subtitles than with dual subtitles?

- 2 Will viewing times differ between the tools?
- 3 Will comprehension differ between the tools?
- 4 Will enjoyability differ between the tools?
- 5 Which of the features of Smart Subtitles will users find helpful and actually end up using?

3.3.4 Study Procedure

Each user was randomly assigned to either of 2 possible groups, depending on which of the tools they would use to watch the clip. The tool for Smart Subtitles was the video player we described above. The tool for parallel subtitles was KMPlayer [7], a video player which is recommended on various language-learning forums for watching videos [9]. Then, we followed the following procedure:

- 1 Fill out a questionnaire asking about prior exposure to Chinese
- 2 Explain to users that they would be shown 2 different clips and would be tested on vocabulary after watching each of them
- 3
 - a Group 1: A 1-minute phase where participants were shown the controls for KMPlayer (pause, seek), then they viewed the first 5-minute video clip with parallel subtitles
 - b Group 2: A 1-minute phase where participants were shown the controls for Smart Subtitles (hover for definitions, show phrase-level translation, click/scroll to navigate dialogs), then they viewed the first 5-minute video clip with Smart Subtitles
- 4 Complete a vocabulary quiz asking them the definitions of 18 words which had appeared in the video clip. [See Appendix 1]
- 5 Answer the following set of qualitative questions about the video viewing experience, with a rating between 1-7:
 - a How easy did you find it to learn new words while watching this video?
 - b How well did you understand this video?
 - c How enjoyable did you find the experience of watching this video with this tool?
- 6 Write a summary describing the clip that you just watched

- a Group 1: A 1-minute phase where participants were shown the controls for Smart Subtitles (hover for definitions, show phrase-level translation, click/scroll to navigate dialogs), then they viewed the second 5-minute video clip with Smart Subtitles
 - b Group 2: A 1-minute phase where participants were shown the controls for KMPlayer (pause, seek), then they viewed the second 5-minute video clip with parallel subtitles
- 8 Complete a vocabulary quiz asking them the definitions of 18 words which had appeared in the video clip. [See Appendix 1]
- 9 Answer the following set of qualitative questions about the video viewing experience, with a rating between 1-7:
- a How easy did you find it to learn new words while watching this video?
 - b How well did you understand this video?
 - c How enjoyable did you find the experience of watching this video with this tool?
- 10 Write a summary describing the clip that you just watched

3.3.5 Results

The answers we concluded from our study were:

- 1 Learners learn more vocabulary with Smart Subtitles than using dual subtitles (approximately 2-3 times more on average)
- 2 Viewing times do not significantly differ between Smart Subtitles and dual subtitles
- 3 Viewers' perceived comprehension of the clips do not differ significantly Smart Subtitles and dual subtitles
- 4 Watching the video clip with Smart Subtitles was considered just as enjoyable as with dual subtitles
- 5 Users made extensive use of both the word-level translations and the dialog-navigation features of Smart Subtitles, and described these as helpful.

We will go into more detail on these individual results below.

Vocabulary Learning

We evaluated vocabulary learning after each of the quizzes via an 18-question free-response vocabulary quiz, with two types of questions. One type of question, shown in *Figure 3.29*, provided a word that had appeared in the video clip, and asked participants to provide the definition for it. The other type of question, shown in *Figure 3.30*, provided a word that had appeared in the video clip, as well as the context in which it had appeared in, and asked participants to provide the definition for it.

For both types of questions, we additionally asked the participant to self-report whether they had known the meaning of the word before watching the video, so that we could determine whether it was a newly learned word, or if they had previously learned it from some external source. This self-reporting mechanism is commonly used in vocabulary-learning evaluations for second-language learning [12].

12)

What does the word 数学 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

Figure 3.29: One type of question asked participants to provide a definition for a word presented in isolation.

2)

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 before watching this video?

Figure 3.30: One type of question asked participants to provide a definition for a word, and showed the context in which it had appeared in the video.

Since the vocabulary quiz answers were done in free-response format, a third-party native Chinese speaker was asked to mark the learners' quiz answers as being either correct or incorrect. The grader was blind as to which condition or which learner the answer was coming from.

As shown in *Figure 3.31*, both the average number of questions which were correctly answered, as well as the number of new words learned, was greater with Smart Subtitles than with dual

subtitles. We measured the number of new words learned as the number of correctly answered questions, excluding those for which they marked that they had previously known the word. 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.

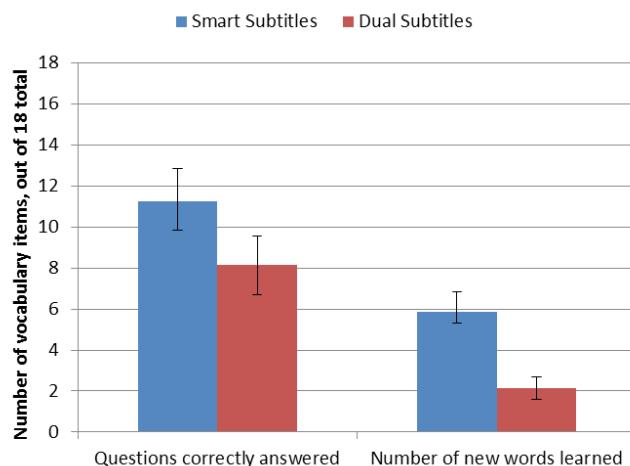


Figure 3.31: Number of vocabulary quiz questions correctly answered in the Smart Subtitles and Dual Subtitles conditions, shown with Standard Error bars.

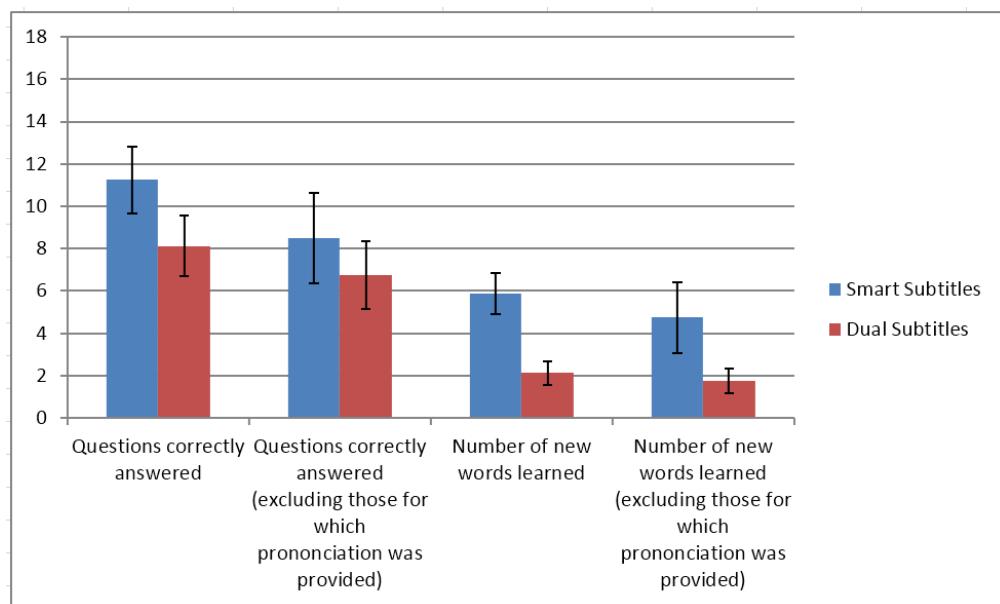


Figure 3.32: Number of vocabulary quiz questions correctly answered, with and without pronunciations provided, shown with Standard Error bars

Although we did not evaluate pronunciation directly, Smart Subtitles' display of pinyin appeared to bring additional attention towards the vocabulary pronunciations. In our vocabulary quizzes, we gave the participants a synthesized pronunciation of the word, in the event that they did not recognize the Chinese characters. We opted to provide a synthesized pronunciation, as opposed to the pinyin directly, as they would not have been exposed to pinyin in the Dual Subtitles condition. This, predictably, allowed participants to correctly define a few additional words. That said, there was a slightly increased level of gain in the Smart Subtitles condition, with an additional 1.1 words correctly answered on average, than in the Dual Subtitles condition, with an additional .3 words correctly answered on average, as shown in *Figure 3.32*.

We attribute this to certain participants focusing more attention on the pinyin, and less on the Chinese characters, in the Smart Subtitles condition. Indeed, one participant remarked during the vocab quiz for Dual Subtitles that she recognized some of the novel words only visually and did not recall their pronunciations. We unfortunately did not ask participants to provide pronunciations for words, only definitions, so we cannot determine if this was a consistent trend.

Viewing Times

As shown in *Figure 3.33*, viewing times did not differ significantly between either of the two 5-minute clips, or between the tools. Viewing times were between 10-12 minutes for each clip, in either condition. Surprisingly, the average viewing times with Smart Subtitles was actually slightly less than with dual subtitles, which is likely due to the dialog-based navigation features. Indeed, during the user study, we observed that users of Smart Subtitles would often review the vocabulary in the preceding few lines of the video clip by utilizing the interactive transcript, whereas users of Dual Subtitles would often over-seek backwards when reviewing, and would lose some time as they waited for the subtitle to appear.

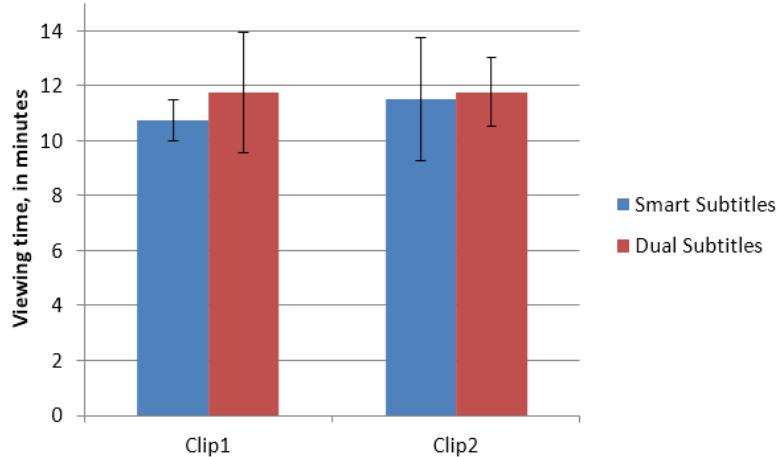


Figure 3.33: Viewing times with Smart Subtitles and Dual Subtitles on each of the clips.

Comprehension and Enjoyment

We evaluated the comprehension and enjoyment aspects of the video viewing tools using qualitative questions, which asked participants to rate, on a scale of 1 through 7:

- How easy did you find it to learn new words while watching this video?
- How well did you understand this video?
- How enjoyable did you find the experience of watching this video with this tool?

As shown in *Figure 3.34*, responses indicated that learners considered it easier to learn new words with Smart Subtitles, ($t=3.76$, $df=7$, $p < 0.005$), and rated their understanding of the videos as similar in both cases. The viewing experience with Smart Subtitles was rated to be slightly more enjoyable on average ($t=1.90$, $df=7$, $p=0.08$).

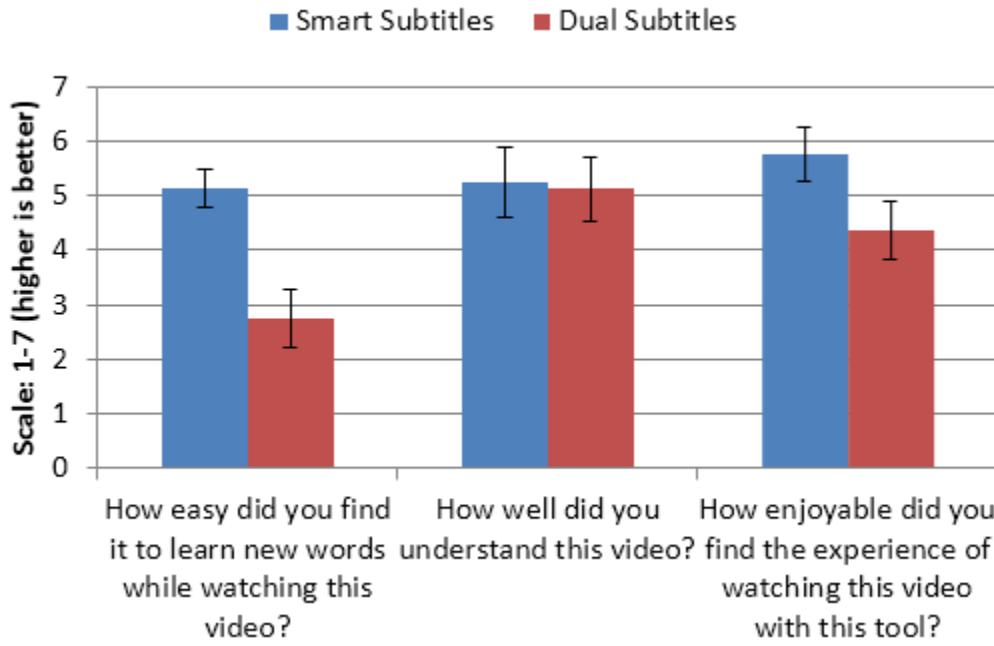


Figure 3.34: Participants' responses to qualitative questions

In addition to the numeric feedback, we also asked participants to write a summary of the video clip they had seen, as well as provide feedback about the watching experience in general. We did not do an evaluation of the summary quality that they had written with third-party experts, though the summaries all appeared to indicate that the viewers understood the video contents in both conditions. The written feedback that participants wrote indicated that they found most of the interface features to be helpful, and would be interested in using Smart Subtitles again. Here is, for example, an anecdote describing the navigation features:

Yes! This was much better than the other tool. It was very useful being able to skip to specific words and sentences, preview the sentences coming up, look up definitions of specific words (with ranked meanings – one meaning often isn't enough), have pinying, etc. 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. With normal subtitling, it's hard to avoid just looking at the English subtitles, even if I can understand it in the foreign language. This also helped when the summation of specific words did not necessarily add up to the actual meaning

The tone coloring feature in particular was not received as well; the only comment that was received was by one participant who described it as distracting. This would suggest that we may

wish to simply remove this feature, or make the tones more salient using another means (perhaps using tone numbers, which are more visually apparent than tone marks):

The tone coloring was interesting, but I actually found it a bit distracting. It seemed like I had a lot of colors going on when I didn't really need the tones color-coordinated. However, I think it's useful to have the tones communicated somehow.

Another user suggested that the interface should perhaps add features to make grammar learning easier (which partially motivated our grammar learning interface described in later chapters):

Absolutely. Something that would be awesome is an additional layer of annotation that indicates template patterns. A lot of the new things I learned watching this weren't new words, per say, but new ways and patterns of using words that I hadn't realized you could do.

Feature Usage During User Studies

During our user studies, we instrumented the interface so that it would record actions such as dialog navigation, mousing over to reveal vocabulary definitions, and clicking to reveal full phrase-level translations.

The degree of feature usage varied widely across participants, though on average, the hover feature was used on 75% of dialog lines (standard deviation 0.22), while the phrase-level translation feature was used on 28% of dialog lines (standard deviation 0.15). Thus, users were using the features as we intended them: relying primarily on hovering to define unfamiliar words, and falling back on phrase-level translations only when word-level definitions were insufficient.

Chapter 4: Grammar Visualization

Our Smart Subtitles system illustrated that word-level translations can oftentimes be sufficient for intermediate-level language learners to comprehend videos. Unfortunately, Smart Subtitles are much less effective with beginning language learners, as these learners do not yet know enough grammar to make sense of a sentence given only word-level translations, and are therefore forced to fall back on sentence-level translations. We aimed to develop a way to scaffold beginning learners so they can continue to comprehend the dialog without knowing the grammar, while helping them learn the grammar. We realized that the resulting grammar visualization could be useful in many domains other than learning during video viewing, so we will describe and analyze it as a standalone system.

4.1 User Interface

Our interface visualizes the syntactic structure of sentences, showing phrase-level translations at each level of the hierarchy, as shown in *Figure 4.1*. Like the word-level glosses generated by services such as Duolingo [33], the visualization can be generated for any sentence; however, the user is still able to clearly see the general grammatical rules that govern the language, regardless of how complex the sentence is.

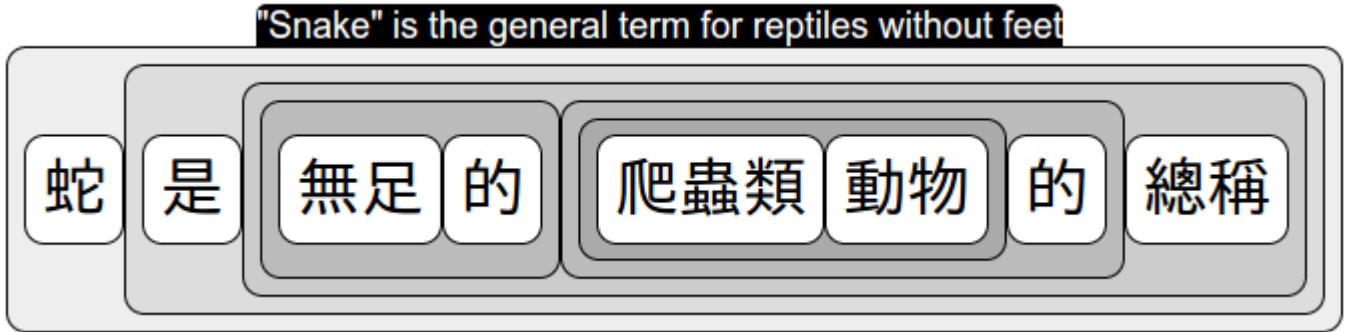


Figure 4.1: The grammar visualization displayed for a sentence.

Our grammar visualization is interactive, and changes according to the particular phrase of the sentence that the user is hovered over, as shown in *Figure 4.2*. The visualization will define the phrase that is hovered over, but will also define surrounding phrases to help the users understand the context which a particular phrase occurs in.



Figure 4.2: When the user hovers over the phrase, the visualization will define both that portion as well as the surrounding phrases.

Our grammar visualization also pronounces the phrase that the user is currently hovered over, to help the user remember the pronunciation of words and phrases that occur in the sentence. In preliminary tests, this feature had positive reception and appeared to increase engagement with the system, with the users often repeating the pronounced phrases upon hearing them. This pronunciation feature, however, was not enabled during our user evaluations on Mechanical Turk, as our evaluations were not focused on pronunciation learning.

Although previous iterations of our grammar visualization included colorization of phrases based on their parts of speech (nouns would be colored green, verbs colored red, etc) to attract attention to the key parts of the sentence, this feature was disabled as users found the colorization distracting.

4.1.1 Applications: Manga Viewer

Because our interface takes up only a little additional space compared to the text itself, it can be embedded into many possible applications where we view foreign language text. For example, we have developed a manga viewer which incorporates this visualization, shown in *Figure 4.3*. By clicking on any bubble in the manga, it will display this visualization for the text in that bubble. Because our grammar visualization system also includes voice synthesis, this application helps learners not only with comprehending the manga and learning vocabulary while reading it, but also with learning the pronunciation of the words.



Figure 4.3: Our foreign-language manga viewer.

4.1.2 Applications: Reading Assistant

Another application of the grammar visualizer is as a reading assistant for foreign-language webpages, shown in *Figure 4.4*. This is useful for when users are reading a foreign-language webpage, and come across a sentence which they do not understand, and would like to see a grammar visualization for. This takes the form of a browser extension which allows users to read a webpage, and click on any sentence of text on the page. It will then display the grammar visualization for that particular sentence.



Figure 4.4: Grammar visualization browser extension shows users the grammar visualization upon clicking on a foreign-language sentence in the webpage.

4.2 Implementation

The grammar visualization requires two pieces of information to generate:

- 1 The phrase-constituency structure of the sentence, which is obtainable from a parse tree
- 2 Translations for each phrase, into the viewer's native language

The feature to pronounce phrases upon mouse-over additionally makes use of voice synthesis.

4.2.1 Obtaining Phrase Structures

The first step in obtaining a parse tree for a sentence is to determine what the particular words in it are. This is done via the process of tokenization, which was described in Section 3.2.3.

Once we have a list of tokens, we can then parse the sentence.

Parse trees can be obtained using a piece of software known as a *parser*. There are two main types of parsers available: *constituency parsers*, and *dependency parsers*. These differ by the type of parse tree that they output.

A *constituency parse*, output by a constituency parser, describes the grammatical class of each of the phrases in the sentence. For example, if we parsed the sentence “John hit the ball” with a constituency parser, it would tell us that “ball” is a noun, “the ball” is a noun phrase, “hit the ball” is a verb phrase, and “John hit the ball” is a sentence. This is generally expressed in computers as an s-expression, where the term immediately following the parenthesis is an abbreviation for the type of phrase, such as “S” for sentence and “VP” for verb phrase. The remainder is the set of word or sub-phrases that this phrase consists of.

(S (N John) (VP (V hit) (NP (Det the) (N ball))))

Alternatively, this can be visualized in a tree structure, shown in *Figure 4.5*.

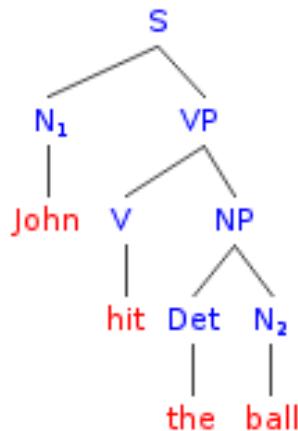


Figure 4.5: A constituency parse for the sentence “John hit the ball”

Thus, coming up with the phrases from a constituency parse is straightforward: they are simply the terminals at each branch of the tree.

In our implementation, we use the Berkeley constituency parser [16] to obtain the subphrases for sentences in English, French, German, and Chinese.

The other type of parser is a dependency parser, which generates a *dependency parse*. A dependency parse lists the relationships between words. For example, in “John hit the ball”, the word “hit” is considered to be the head of this sentence, since it is the main verb, and it depends on “John”, since John is the subject, and on “ball”, since the ball is the object. This can be

visualized with a diagram where the head words point to their dependents, with the arrows indicating the dependency relationship, as shown in *Figure 4.6*.

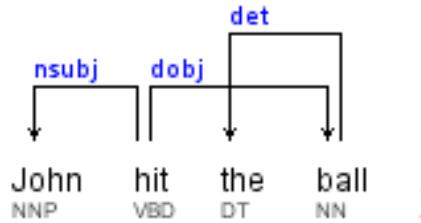


Figure 4.6: A dependency parse for the sentence “John hit the ball”

We can alternatively visualize this in a tree structure, where we put the head constituents at the root, and their dependents as children, as shown in *Figure 4.7*.

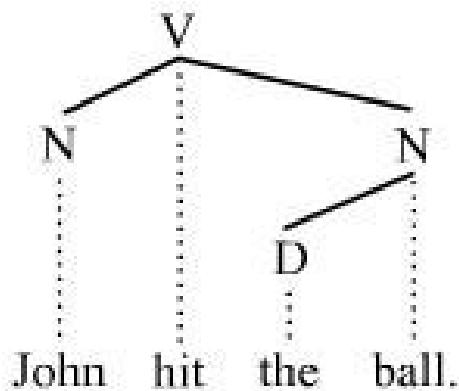


Figure 4.7: An alternative visualization for dependency parses.

As we can see from here, we can derive the subphrases in a sentence from a dependency parse: we consider the dependency tree, and treat each head and its dependents as a subphrase. In our implementation, we use the KNP parser [18] to obtain a dependency parse for Japanese sentences, for use in determining the subphrases. We use a dependency parser instead of a constituency parser one for Japanese, because only dependency parsers are available for Japanese.

4.2.2 Phrase Translation

At this point, we have obtained the hierarchy of subphrases that a given sentence consists of. Now, for each of them, we need to obtain translations of each of these subphrases to the user's

native language. We can obtain these via machine translation. Our implementation has the option of using either Google's machine translation service, or Microsoft's service. Additionally, we look up terminal words to see whether they are present in a bilingual dictionary, so that we can display alternative definitions and romanizations for them if they are available. For languages such as Japanese, this process also involves morphological analysis and stemming, which was described earlier.

4.2.3 Voice Synthesis

The voice synthesis component of our grammar visualization - which tells users how to pronounce the particular phrase in the sentence that is hovered over - is implemented using the HTML5 audio tag. The voice synthesis engine used is the same one that is available for usage in Google Translate. Because Google's voice synthesis system imposes various usage limitations, such as limiting the frequency of synthesis requests and limiting the length of the text that can be synthesized, we overcome these with server-side audio caching workarounds. The voice synthesis component was not enabled during our tests on Mechanical Turk.

4.3 User Study

Our user study for the grammar visualization wished to compare its effectiveness in helping non-bilingual people translate foreign-language sentences, to existing ways that foreign-language sentences are presented to people. Specifically, we were comparing the grammar-structure visualization against simply presenting users with a machine translation, and to a visualization that provided word-level translations in addition to a machine translation.

4.3.1 Participants

Our study participants were Mechanical Turk [36] workers from the United States, who had completed at least 1000 HITs with at least 98% approval rating. As we did not have any language qualification requirement, and our users were from the United States, these were presumably all monolingual English speakers.

4.3.2 Materials

Users were asked to translate 30 sentences from Chinese to English. These are listed with reference English translations in Appendix 2. These sentences were randomly sampled from the Tatoeba corpus of Chinese sentences [34]. We chose this particular source of sentences, as the sentences were all written so that they could be understood without additional context. We restricted sentences to those that were 21 Chinese characters in length, to ensure that there was roughly uniform difficulty in each task. This particular length was selected to ensure that the sentences would be sufficiently complex that the machine translations would require postediting, while not being so long that it would take the users extraordinarily long times to make the translations.

4.3.3 Conditions

Our experimental conditions compare three types of visualizations for displaying sentences for postediting. In the *structural postedit* condition, the visualization shows the full grammatical structure, as shown in *Figure 4.8*.

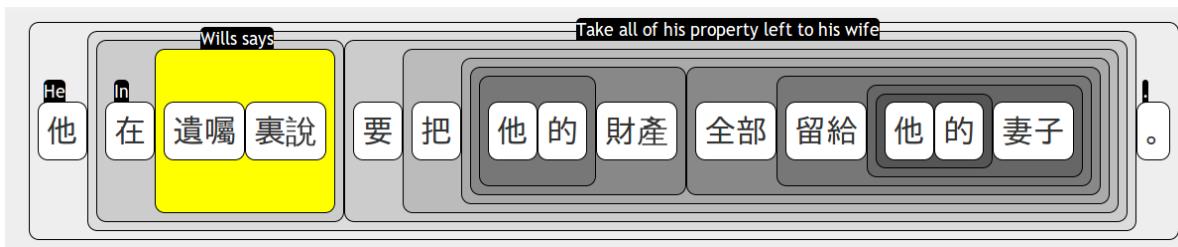


Figure 4.8: Structural postedit condition

In the *term-level postedit* condition, the visualization presents the same interface, except it shows the outermost and innermost levels, as shown in *Figure 4.9*. In other words, it provides the user with only the full-sentence translation and word-level translations. This is intended to mimic the word-level information that is provided to users by tools such as Rikaikun [31] or services such as Duolingo [33]:



Figure 4.9: Term-level postedit condition

In the *simple postedit* condition, the visualization simply shows the original sentence, and presents a machine translation below it, as shown in *Figure 4.10*.



他在遺囑裏說要把他的財產全部留給他的妻子。
His will that all his property should be left to his wife.

Figure 4.10: Simple postedit condition

4.3.4 Procedure

In each HIT, the participants are first presented with a description of the task, as well as an example illustrating each of the types of visualizations that they will be shown, as shown in *Figure 4.11*.

Next, the user clicks on an individual sentence to show it and begin translating. The 3 sentences are each presented with a different one of the 3 possible visualizations. The user is asked to provide a translation for each one of the sentences, as shown in *Figure 4.12*. Our system times the period that the user has the visualization for a given sentence open, in order to determine whether there is any difference in time used for translating under each condition. After translating each of the 3 sentences, the user is finally given the option to provide free-form feedback about the task, and submit it.

We posted 100 of these HITs, covering 30 sentences, on Mechanical Turk. Workers were paid 50 cents for completing each task. We obtained 10 translations for each sentence through the process.

Afterwards, we used a pair of professional translators on oDesk [37] to evaluate the quality of these translations. The translation evaluators were asked to rate each translation, on a scale of 1 (worst) to 9 (best) based on “fluency” (how fluent the English in the translation is), and “fidelity” (how well the translation preserves the original sentence’s meaning). This translation quality evaluation scale is the one recommended by the Automatic Language Processing Advisory Committee [38]. In our case, we were interested particularly in the fidelity measure, since it reflects how well the user understood the sentence.

Fix Machine-translation Errors

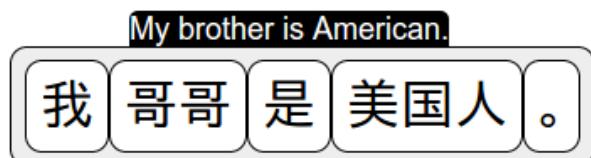
You will be shown 3 sentences in a foreign language. Each sentence will be shown along with a (possibly incorrect) machine-generated translation to English. Correct the English translation, and enter it into the textbox. Note that you do NOT need to know the foreign language to do this task (we are just trying to see how well you understand the machine-generated translation).

The 3 different sentences will each be shown in a different way. One sentence will be shown with the original (foreign-language) sentence, followed by the machine translation, as shown below:

我哥哥是美国人。

My brother is American.

Another sentence will show the machine translation for the sentence if you hover the outside, and translations for the individual words if you hover over them, as shown below:



Another sentence will show the machine translation for the sentence, and translations for any of the constituent phrases if you hover over their outlines, as shown below:



▶ Sentence 1

▶ Sentence 2

▶ Sentence 3

▶ Feedback

Submit

Figure 4.11: Part of the HIT shown to the user, giving instructions and examples prior to starting translation of the individual sentences.



Figure 4.12: Translation interface shown to users for translating an individual sentence.

4.3.5 Results

Post-editing Quality

The 2 oDesk evaluators largely agreed on the translation fidelity scores they assigned to the translations generated by the study participants; the Pearson Correlation Coefficient [35] between their ratings was 0.7.

As shown in *Figure 4.13*, the average translation fidelity scores showed a significant improvement in any of the postediting conditions relative to machine translation. Unfortunately, there was no significant improvement in the fidelity scores of translations generated with the structural visualization, compared to the other two postediting conditions. Additionally, the quality of the postedited translations were still significantly below the reference translations in quality.

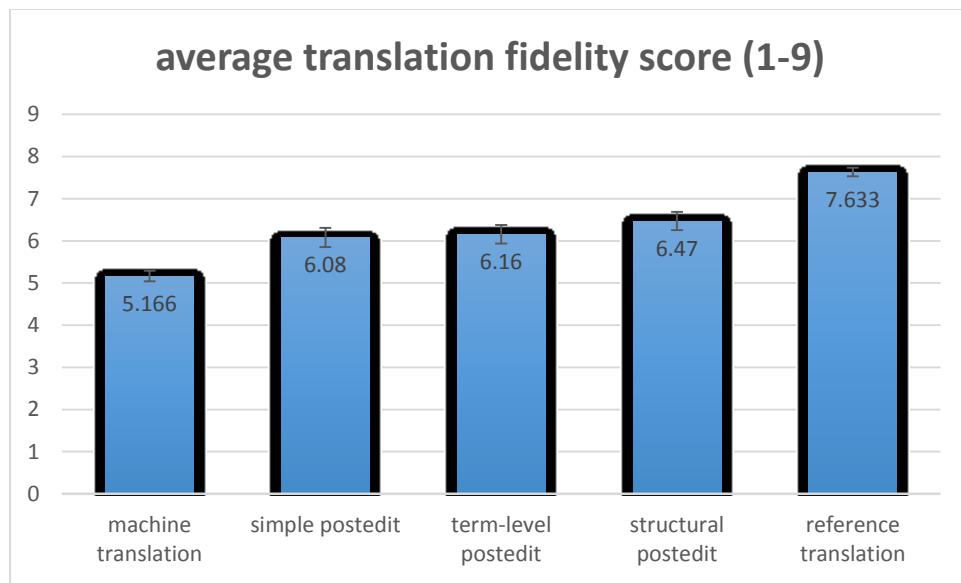


Figure 4.13: Average translation fidelity scores in each condition, with standard error bars.

However, as shown in *Figure 4.14*, if we take the best translation generated for each sentence in each condition, we find that the best translation generated in the structural and term-level postedit conditions is not significantly different from the reference translation.

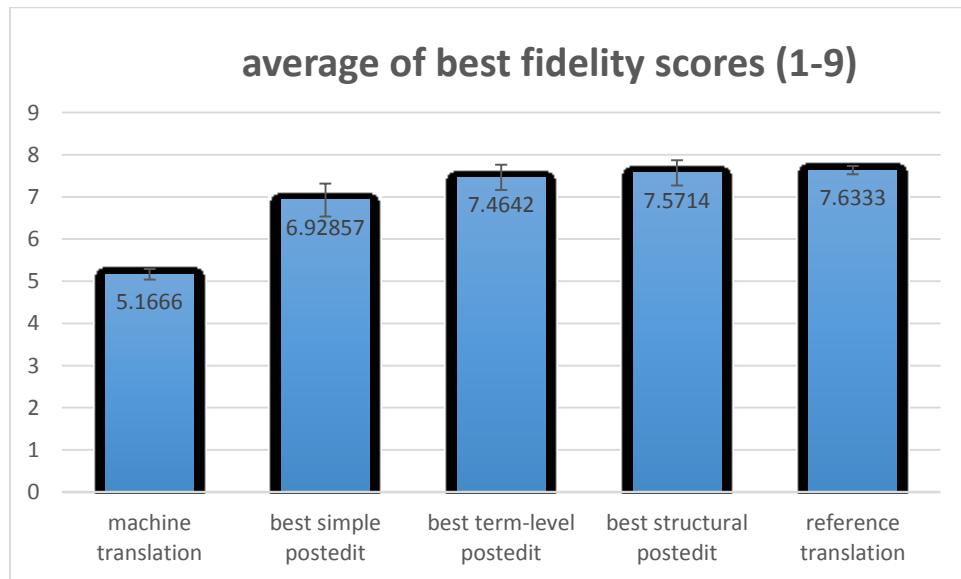


Figure 4.14: Averages of the best translation fidelity scores for each sentence, with standard error bars.

Additionally, there were a number of sentences where the machine translations were already sufficiently high quality that postediting would have little benefit in improving the translation fidelity. From the perspective of our visualization as a means of comprehension assistance,

these would be the sentences that a reader would likely be able to understand without needing our visualization. In our case, there were 6 sentences where the machine translation fidelity was greater than or equal to 6 on a 9-point scale. Therefore, we looked at what happens if we exclude these sentences from our analysis, restricting our attention to those sentences which are most in need of postediting. As shown in *Figure 4.15*, the advantage of presenting the grammatical-structure visualization relative to simply presenting a machine-translation becomes more profound:

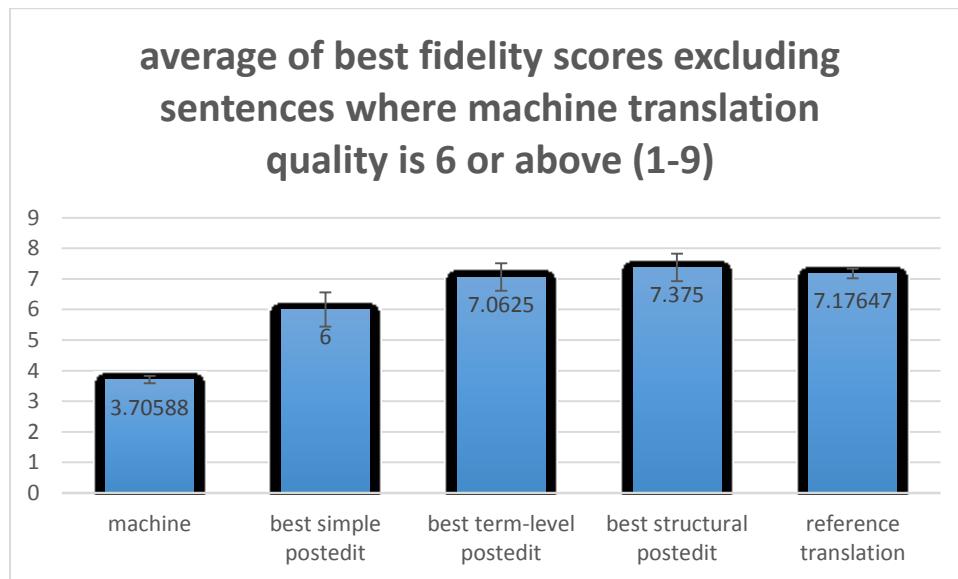


Figure 4.15: Averages of best translation fidelity scores on sentences where machine translation performed poorly, with standard error bars.

As shown in *Figure 4.16*, there is an increase in the average amount of time a user spends when translating with either the structural visualization or the term-level visualization, relative to simply postediting the raw machine translation output. Interestingly, users did not spend additional time in the structural postedit condition, relative to the term-level condition. This suggests that users didn't fully explore the additional information presented by the grammatical structure visualization in the structural postedit condition.

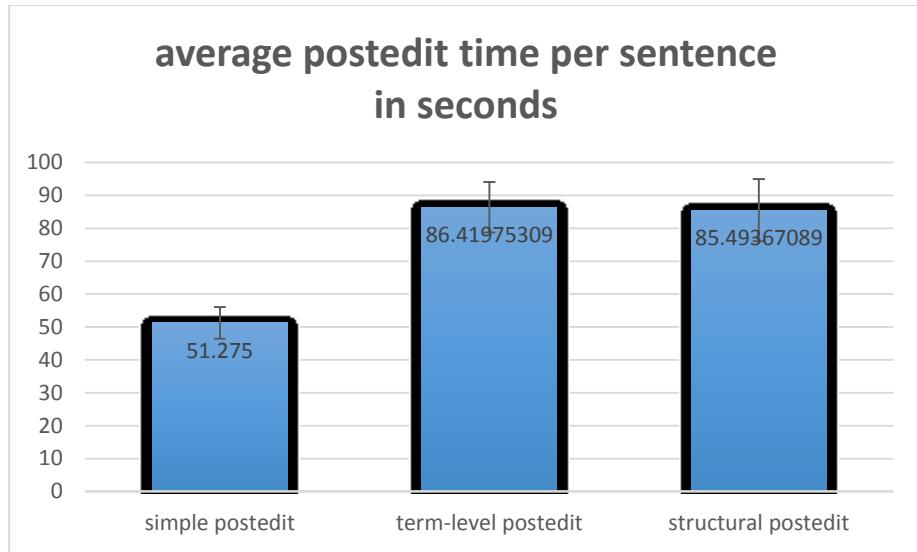


Figure 4.16: Average translation postediting times per sentence, in seconds, with standard error bars.

Our results suggest that the structural grammar visualization can help users gain a better understanding of the sentence, relative to standard postediting procedures, particularly in situations where the machine translation performed poorly.

Qualitative Feedback

The feedback users left suggests that they found the structural visualizations helpful. The most common feedback provided was that sentences lacking the structural visualization did not provide enough information, suggesting that they found the information provided by the structural visualizations helpful for translation. For example:

can't possibly translate this without the "hover over" version.. it's too messed up!

Many also commented that they found the experience of exploring the structure of the sentences with our visualization to be interesting or enjoyable. For example:

Interesting task, thank you. Each sentence could be written many different ways; I tried to keep most of the original words and structure, where possible.

Chapter 5: Discussion

We have developed and evaluated a pair of systems for helping users learn language and comprehend authentic foreign-language materials.

Our first system, Smart Subtitles, aims to improve vocabulary learning while foreign-language learners are watching foreign-language videos. It does so by providing features to compensate for learners' vocabulary deficiencies, and by providing navigation features to help the learner review the video to ensure they learn the vocabulary while still understanding the story. Our user study with intermediate-level learners found an increase in vocabulary acquisition and no decrease in comprehension or enjoyability, relative to dual Chinese-English subtitles.

Additionally, we found that users were able to comprehend the dialog while needing to refer to full-sentence translations on only 25% of the dialogs, suggesting that this system will also be useful for assisting in video comprehension when no English-language subtitle is available.

Though we did not compare against the GliFlix approach of inserting common foreign words into the English subtitle [30], we believe our approach has the advantage that it focuses attention on the foreign-language transcript and vocabulary, which we believe may have positive effects on factors beyond simply vocabulary acquisition, such as pronunciation and grammar (that said, our user study remained focused on vocabulary). Additionally, Smart Subtitles can be generated

automatically simply from a transcript, rather than requiring someone to have generated a manually translated subtitle beforehand.

While our Smart Subtitles user study with intermediate-level learners achieved our desired results of an increase in vocabulary acquisition and no decrease in comprehension or enjoyability, we believe we would be unable to reproduce these results with beginning learners. Specifically, beginning learners would not have the grammatical knowledge to understand the dialog if they were simply given the word-level translations that Smart Subtitles provides. Therefore, we developed a grammar visualization to provide a scaffold for learners who still need to learn the grammar of the language, while helping them learn grammar and vocabulary by reading arbitrary foreign-language sentences.

We believe that our grammar visualization has applications beyond grammar learning - specifically, we believe it also improves comprehension of foreign-language sentences. Therefore, our user study for the grammar visualization tested it as a machine-translation postediting system, to see whether users produce better translations when using our grammar visualization. There already do exist certain visualizations that are optimized towards helping non-bilinguals produce translations, such as the Chinese Room [21], though our system has the benefit that we believe it is more intuitive for novices, as we did not require an extensive training phase, nor do we expect users to be familiar with linguistics terminology. Additionally, it can be embedded essentially anywhere, as opposed to requiring the large amount of screen space used by the visualization for the Chinese Room. While our user study showed no significant improvement in the average fidelity of translations produced by non-bilingual users using our grammar visualization, we did observe a significant improvement if we restrict ourselves to sentences which the machine translation performed poorly on, and if we consider the best of the translations produced among all users. Qualitative feedback by study participants likewise shows that they feel the grammar visualization is enjoyable to use and helpful in comprehending the sentence. This suggests that our grammar visualization may be beneficial in the scenario of a dedicated non-bilingual user striving to comprehend a sentence where machine translation has failed to convey the meaning.

Both of these tools encourage users to interact closely with the foreign-language text to help them comprehend it. Smart Subtitles encourages users to read definitions for unknown words, whereas our grammar visualization encourages them to dig through the phrase-level structure

to better understand its meaning. Placing these tools in the context of foreign-language multimedia - videos in the case of Smart Subtitles, and manga in the case of our grammar visualization - will thus overcome the common issue in translated media, in that the foreign-language learning objectives are sacrificed for the sake of comprehending and enjoying the material. With our tools, we believe that comprehension and enjoyment of the material can be achieved through the process of closely interacting with the foreign text, which will help learners learn the foreign language.

Chapter 6: Future Work and Conclusion

As we have shown, we can build systems that will help learners experience authentic foreign-language content, relying minimally on professional translations, while still allowing users to comprehend, enjoy, and learn from the material. Smart Subtitles have illustrated that intermediate-level learners can enjoy and comprehend videos without requiring that subtitles be displayed by default, and learn more from it than from existing subtitles. Our grammar visualization has illustrated that it can help even users with no experience to the language to make sense of complex foreign-language sentences.

Much work can still be done in the area of incorporating multimedia into learning. In particular, we have thus far focused on (written) vocabulary learning in the Smart Subtitles system, and on comprehension in our grammar visualization system. We believe that multimedia augmentations can also benefit the other aspects of language learning - in particular, pronunciation, listening comprehension ability, and grammar learning. For example, we expect that the voice synthesis integrated into the grammar visualization system should presumably help with the pronunciation aspect in particular, as we observed in preliminary testing with our manga reader that users would often repeat back the synthesized phrases upon hearing them pronounced.

In addition, we can explore additional means of engaging learners while they are consuming multimedia. For example, in the Smart Subtitles system, we observed that in informal, non-lab settings users would gradually make less use of the system's features if they were using it for prolonged periods of time, suggesting that they were growing disengaged. We could potentially detect when users grow disengaged, and strategically embed vocabulary quizzes into the video stream to ensure that they continue making active efforts towards learning while using the system. Alternatively, we can personalize the learning experience to detect harder portions of the dialog and provide additional scaffolding for lower-level learners - perhaps slowing down the rate of dialog, or showing the English translations for certain phrases by default. This reduces the amount of required interaction from users, helping to avoid fatigue over prolonged usage.

In addition to the systems and media we have focused on - videos and manga - additional systems can potentially be built to assist learners with passion for different forms of multimedia. For example, Smart Subtitles can be easily used with songs, provided that we have obtained the timing information ahead of time. However, songs have the additional property that people often enjoy singing them - therefore, we could potentially explore augmenting traditional games such as karaoke to support language learning.

This work can potentially lead to a future where people can learn foreign languages more enjoyably by being immersed and enjoying the culture of foreign countries, in the form of their multimedia, without requiring dedicated effort towards making the material education-friendly or even fully translating it.

Bibliography

- [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] Danan, Martine. "Reversed subtitling and dual coding theory: New directions for foreign language instruction." *Language Learning* 42.4 (1992): 497-527.
- [5] d'Ydewalle, Géry. "Foreign-language acquisition by watching subtitled television programs." *Journal of Foreign Language Education and Research* 12 (2002): 59-77.
- [6] Bianchi, Francesca, and Tiziana Ciabattoni. "Captions and Subtitles in EFL Learning: an investigative study in a comprehensive computer environment." (2008).
- [7] KMPMedia. <http://www.kmpmedia.net/>
- [8] Dummitt, Nathan. *Chinese Through Tone & Color*. Hippocrene Books, 2008.
- [9] Using KMPlayer to watch a movie with 2 subtitles at the same time
<http://boardgamegeek.com/blogpost/8836/using-kmplayer-to-watch-a-movie-with-2-subtitles-a>
- [10] Learn Chinese with Yangyang 001 <http://www.youtube.com/watch?v=zNoOdNvdZlg>
- [11] Jay Chou 周杰伦- Silence 安靜 English + Pinyin Karaoke Subs
<http://www.youtube.com/watch?v=QvN898LaTTI>
- [12] Wesche, Marjorie, and T. Sima Paribakht. "Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth." *Canadian Modern Language Review* 53.1 (1996): 13-40.
- [13] WebVTT: The Web Video Text Tracks Format. <http://dev.w3.org/html5/webvtt/>

- [14] Huihsin Tseng, Pichuan Chang, Galen Andrew, Daniel Jurafsky and Christopher Manning. 2005. *A Conditional Random Field Word Segmenter*. In Fourth SIGHAN Workshop on Chinese Language Processing.
- [15] Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python*. O'Reilly Media, 2009.
- [16] "Learning Accurate, Compact, and Interpretable Tree Annotation" Slav Petrov, Leon Barrett, Romain Thibaux and Dan Klein in COLING-ACL 2006
- [17] S. Kurohashi, T. Nakamura, Y. Matsumoto and M. Nagao, Improvements of Japanese morphological analyzer JUMAN," In Proceedings of Int. Workshop on Sharable Natural Language Resources, Nara, Japan, 1994.
- [18] Kurohashi, Sadao, and Makoto Nagao. "Kn parser: Japanese dependency/case structure analyzer." *Proceedings of the Workshop on Sharable Natural Language Resources*. 1994.
- [19] CC-CEDICT. 2013. Available from <http://cc-cedict.org/wiki/>
- [20] WWWJDIC. 2013. Available from <http://wwwjdic.org>
- [21] Albrecht, Joshua S., Rebecca Hwa, and G. Elisabeta Marai. "Correcting automatic translations through collaborations between MT and monolingual target-language users." *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009.
- [22] Xiaoyi Ma. 2005. *Chinese English News Magazine Parallel Text*. Linguistic Data Consortium, Philadelphia
- [23] CC-CEDICT. 2012. Available from <http://cc-cedict.org/wiki/>
- [24] Bansal, Mohit, John DeNero, and Dekang Lin. 2012. "Unsupervised Translation Sense Clustering." *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*
- [25] Kristina Toutanova and Christopher D. Manning. 2000. *Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger*. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, pp. 63-70.
- [26] SenseEval-3 Data. ACL-SIGLEX. <http://www.senseval.org/senseval3>
- [27] D. Albanese, R. Visintainer, S. Merler, S. Riccadonna, G. Jurman, C. Furlanello. *mipy: Machine Learning Python*, 2012. [arXiv:1202.6548](https://arxiv.org/abs/1202.6548) [bib]
- [28] DeNeefe, Steve, Kevin Knight, and Hayward H. Chan. "Interactively exploring a machine translation model." *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions*. Association for Computational Linguistics, 2005.
- [29] Callison-Burch, Chris. "Linear B system description for the 2005 NIST MT evaluation exercise." *Proceedings of the NIST 2005 Machine Translation Evaluation Workshop*. 2005.
- [30] Sakunkoo, Nathan, and Pattie Sakunkoo. "GliFlix: Using Movie Subtitles for Language Learning. *UIST 2009 Companion (demo paper)*, Vancouver, Canada, Oct 2009.
- [31] Chrome Web Store - Rikaikun.
<https://chrome.google.com/webstore/detail/rikaikun/jipdnfibhldikgcjhfnomkfpcebammhp?hl=en>
- [32] Yu, Chen-Hsiang, and Robert C. Miller. "Enhancing web page skimmability." *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*. ACM, 2012.
- [33] Duolingo. <http://duolingo.com/>

- [34] Tatoeba project. <http://tatoeba.org>
- [35] Kornbrot, Diana. "Pearson Product Moment Correlation." *Encyclopedia of Statistics in Behavioral Science* (2005).
- [36] Amazon Mechanical Turk. <https://www.mturk.com>
- [37] oDesk. <https://www.odesk.com/>
- [38] ALPAC (1966) "Languages and machines: computers in translation and linguistics". A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council, 1966. (Publication 1416.)
- [39] Geza Kovacs, Joy Chen, Xinyi Zhang. Translation Sense Disambiguation on Chinese-English. <http://geza.csail.mit.edu/translation-sense-disambiguation.pdf>
- [40] Ripping subtitles from video files using SubRip. <http://zuggy.wz.cz/guides/video.htm>
- [41] Wiktionary. <http://www.wiktionary.org/>
- [42] Microsoft OneNote. <http://office.microsoft.com/en-us/onenote/>
- [43] OneNote for Developers. <http://msdn.microsoft.com/en-us/office/aa905452.aspx>
- [44] Smith, Ray. "An overview of the Tesseract OCR engine." *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on.* Vol. 2. IEEE, 2007.
- [45] C. Harris and M. Stephens (1988). "A combined corner and edge detector". Proceedings of the 4th Alvey Vision Conference. pp. 147–151.
- [46] OpenCV. <http://opencv.org/>
- [47] H. Samet and M. Tamminen (1988). "Efficient Component Labeling of Images of Arbitrary Dimension Represented by Linear Bintrees". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TIEEE Trans. Pattern Anal. Mach. Intell.)* 10: 579. doi:10.1109/34.3918.

Appendices

Appendix 1: Questionnaire

Foreign Language Video Viewing Study

Questionnaire

I. Basic User Information

- Are you currently enrolled in a Chinese class? If yes, which one? _____
- How long have you been studying Chinese? _____
- Other than classes, where have you learned Chinese from? _____
- Do you watch Chinese-language videos on your own (when not required by a class)?
 Yes
 No

If you answered yes:

What type of Chinese-language videos? [Check all that apply]

- Movies
- Talk shows
- Dramas
- Other: _____

What type of subtitles do you usually watch Chinese-language videos with?

- No subtitles
- Chinese subtitles
- English subtitles
- Chinese+English dual subtitles

II. General Questions for Each Video

Q1: How easy did you find it to learn new words while watching this video?

Very Difficult	Normal	Very Easy				
1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

Q2: How well did you understand this video?

Do Not Understand	Normal	Completely Understand				
1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

Q3: How enjoyable did you find the experience of watching this video with this tool?

Very Unenjoyable	Normal	Very Enjoyable				
1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	7 <input type="radio"/>

Q4: Please describe what this video clip was about.

Q5: If given the option, would you use this tool on your own? Do you have any other comments?

Q1: How easy did you find it to learn new words while watching this video?

Very Difficult	Normal			Very Easy		
1	2	3	4	5	6	7
<input type="radio"/>						

Q2: How well did you understand this video?

Do Not Understand	Normal			Completely Understand		
1	2	3	4	5	6	7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				

Q3: How enjoyable did you find the experience of watching this video with this tool?

Very Unenjoyable	Normal			Very Enjoyable		
1	2	3	4	5	6	7
<input type="radio"/>						

Q4: Please describe what this video clip was about.

Q5: If given the option, would you use this tool on your own? Do you have any other comments?

我是老师

1)

In the following sentence, what does the word 为止 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

2)

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 before watching this video?

3)

In the following sentence, what does the word 当 mean?

我没有资格当老师

Meaning: _____

Did you already know the meaning of this word before watching this video?

4)

In the following sentence, what does the word 绝对 mean?

我绝对不会放弃希望的

Meaning: _____

Did you already know the meaning of this word before watching this video?

5)

In the following sentence, what does the word 放弃 mean?

我绝对不会放弃希望的

Meaning: _____

Did you already know the meaning of this word before watching this video?

6)

In the following sentence, what does the word 希望 mean?

我绝对不会放弃希望的

Meaning: _____

Did you already know the meaning of this word before watching this video?

7)

In the following sentence, what does the word 班长 mean?

班长我们做到哪了

Meaning: _____

Did you already know the meaning of this word before watching this video?

8)

In the following sentence, what does the word 滋味 mean?

有你我才觉得有活的滋味啊

Meaning: _____

Did you already know the meaning of this word before watching this video?

9)

In the following sentence, what does the word 痛 mean?

往这边打才更痛

Meaning: _____

Did you already know the meaning of this word before watching this video?

10)

In the following sentence, what does the word 到底 mean?

到底有没有良心？

Meaning: _____

Did you already know the meaning of this word before watching this video?

11)

What does the word 良心 mean?

到底有没有良心？

Meaning: _____

Did you already know the meaning of this word before watching this video?

12)

What does the word 数学 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

13)

What does the word 本来 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

14)

What does the word 干嘛 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

15)

What does the word 概念 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

16)

What does the word 会长 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

17)

In the following sentence, what does the word 至少 mean?

请了 9 千万的老师至少应该做做学习的样子啊

Meaning: _____

Did you already know the meaning of this word before watching this video?

18)

What does the word 丫头 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

=

1)

In the following sentence, what does the word 敲一下 mean?

不是说敲一下门吗

2) What does the word 毕业 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

3) In the following sentence, what does the word 不管 mean?

我不管别人怎么说

Meaning: _____

Did you already know the meaning of this word before watching this video?

4) What does the word 留学 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

5) In the following sentence, what does the word 情况 mean?

那么你们就在没有老师的情况下学习了吗?

Meaning: _____

Did you already know the meaning of this word before watching this video?

6) In the following sentence, what does the word 尊敬 mean?

尊敬感有什么大不了吗

Meaning: _____

Did you already know the meaning of this word before watching this video?

7) In the context of the following sentence, what does the word 胆敢 mean?

吃了豹子胆敢对我的女儿动手

Meaning: _____

Did you already know the meaning of this word before watching this video?

8) In the context of the following sentence, what does the word 溺爱 mean?

溺爱子女是错误

Meaning: _____

Did you already know the meaning of this word before watching this video?

9) In the context of the following sentence, what does the word 错误 mean?

溺爱子女是错误

Meaning: _____

Did you already know the meaning of this word before watching this video?

10) What does the word 嫁 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

11) In the following sentence, what does the word 话题 mean?

又是这个话题

Meaning: _____

Did you already know the meaning of this word before watching this video?

12) In the following sentence, what does the word 后悔 mean?

不要后悔

Meaning: _____

Did you already know the meaning of this word before watching this video?

13) What does the word 禁止 mean in the context of the following sentence?

从今天开始禁止出入知道了吗

Meaning: _____

Did you already know the meaning of this word before watching this video?

14) What does the word 家伙 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

15) What does the word 蜘蛛 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

16) What does the word 杀死 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

17) What does the word 做梦 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

18) What does the word 苦恼 mean?

Meaning: _____

Did you already know the meaning of this word before watching this video?

Appendix 2: Sentences Used in Grammar Visualization Study

These are the sentences that we used in the grammar visualization study. They are from the Tatoeba corpus of Chinese sentences [34]. They are shown alongside the reference translation to English, as well as the machine translation generated by Google translate.

1)

我上機之前打了電話給兒子，叫他去機場接我。

I phoned my son before boarding the plane, telling him to come to the airport to pick me up.

Son telephoned before me on the plane, told him to go to the airport to pick me up.

2)

他整天都在埋頭埋腦地看小說，連飯也沒有吃。

He has been absorbed in the novel all day without eating.

All day buried buried brain to read the novel, even rice have nothing to eat.

3)

不知道他放假的時候可不可以把結他借給我呢？

I wonder if he can lend me his guitar during the vacation.

Do not know his holiday Can Guitar lend me?

4)

他在遺囑裏說要把他的財產全部留給他的妻子。

He left all his property to his wife in his will.

His will that all his property should be left to his wife.

5)

老師和她愛人帶着參加婚禮的人來到一個公園。

The teacher and her partner took the people attending the wedding to a park.

The teacher and her lovers with the wedding party came to a park.

6)

那位作家用過的傢俱全都存放在這間博物館裏。

The writer's furniture is all shown in this museum.

The writer used furniture are all stored in this museum.

7)

醫生雖然已經竭盡全力，但不久病人還是死了。

All the doctor's efforts were in vain and the man soon died.

The doctor though Every effort has been made, but near the patient or died.

8)

他每個月都會寫一封信給他的父母，風雨不改。

He never fails to write to his parents once a month.

Him to write a letter every month to his parents Fengyubugai.

9)

盤子上盛着一塊雞肉、一塊馬鈴薯和一些青豆。

On the plate was a piece of chicken, a potato and some green peas.

Candy in a plate of chicken, a potato and some green beans.

10)

你去那個超級市場，幾乎甚麼日用品都買得到。

If you go to that supermarket, you can buy most things you use in your daily life.

You go to the supermarket to buy almost anything other daily necessities.

11)

坐公車從這裏到機場只不過是十五分鐘的路程。

It's only a fifteen minute bus ride from here to the airport.

Take a bus from here to the airport is only 15 minutes away.

12)

這就是駝鳥肉比牛肉和豬肉貴兩倍多的原因了。

That's why ostrich meat costs more than twice as much as beef and pork.

This is the ostrich meat is more expensive than beef and pork twice as many reasons.

13)

那間舊屋裏只有一張床，所以我們便輪流睡覺。

The old cottage had only one bed, so we all took turns sleeping in it.

The old house is only one bed, so we take turns to sleep.

14)

他想成為有錢人，所以便從早到晚不斷地工作。

He worked day and night so that he might become rich.

He wanted to become rich, so it from morning to night work.

15)

小李往往失眠是因为他的邻居老是大声地吵架。

Xiaoli often loses sleep because his neighbours are always arguing loudly.

Li often insomnia is because his neighbors always quarrel loudly.

16)

黑暗中突然有一隻狗冒了出來，把我嚇了一跳。

I was shocked by the dog that sprang out of the darkness.

The dark suddenly a dog run out, I was shocked.

17)

從學校回來的路上讓我們來討論你的愛情問題。

Let's discuss your love problems on the way back from school.

On the way back from the school let's talk about your love problems.

18)

直到露西離開了我，我才意識到我是多麼愛她。

It was not until Lucy left me that I realized how much I loved her.

Until Lucy left me, I did not realize how much I love her.

19)

她把她錢包裏僅有的幾枚硬幣都給了那個男孩。

She gave the boy what few coins she had in her purse.

Only a few coins in her purse she gave a boy.

20)

如果在暗中做善事，也總是有一個良好的回報。

Secret gifts are openly rewarded.

If you do good in the dark, there is always a good return.

21)

人有選擇的權利，但是奴隸卻只有服從的份兒。

A man chooses; a slave obeys.

The people have the right to choose, but the slaves in the negotiations but only to obey.

22)

賽義認為數學家自給自足的獨行於一小塊天堂。

Saeb thinks mathematicians are self-contained walking pieces of heaven.

The Seydou that lone mathematician self-sufficiency in a small piece of paradise.

23)

一般來說，你應該為每兩個客人準備一磅牛肉。

As a rule of thumb, you should plan on one pound of beef for every two guests.

In general, you should be prepared for every two guests a pound of beef.

24)

根據我的經驗，掌握法語語法需要一年的時間。

According to my experience, it takes one year to master French grammar.

According to my experience, to master French grammar needs a year's time.

25)

忧愁就像魔鬼一样，它伤害不了不害怕它的人。

Worries are like ghosts - those who aren't afraid of them can't suffer at their hands.
Sorrow, as the devil, it can not hurt not to be afraid of its people.

26)

他用斧头重重的斩了一下，那棵树就倒了下来。

The tree was felled with one hard blow of his ax.
Him with an ax heavy chop that tree down.

27)

声音大声一点，以便让坐在后边的学生能听到。

Turn the volume up so that the students at the back can hear.
Sounds louder, so that sitting behind the students can hear.

28)

我只跟你说，我很快就要辞掉我现在的工作了。

Between you and me, I'm going to quit my present job soon.
I just told you, I soon quit my job now.

29)

比如说，我要到市中心，有什么车子可以坐吗？

For example, if I want to get downtown what bus should I take?
For example, I have to go to the city center, a car can take you?

30)

因为今天明天我休息，所以趁机会快点搬好它。

Because I don't have to go to work today and tomorrow, so I will use the opportunity to quickly finish moving it.

Rest, because today tomorrow, I will take the opportunity to quickly move it well.