Smart Subtitles for Vocabulary Learning

Geza Kovacs
Stanford University
Palo Alto, CA, USA
gkovacs@stanford.edu

Robert C. Miller
MIT CSAIL
Cambridge, MA, USA
rcm@mit.edu

ABSTRACT
Language learners often use subtitled videos to help them learn the language. However, standard subtitles are suboptimal for vocabulary learning, as translations are nonliteral and made at the phrase level, making it hard to find connections between the subtitle text and the words in the video. This paper presents Smart Subtitles, which are interactive subtitles tailored towards vocabulary learning. Smart Subtitles can be automatically generated from common video sources such as subtitled DVDs. They provide features such as vocabulary definitions on hover, and dialog-based video navigation. Our user study shows that students studying Chinese learn over twice as much vocabulary with Smart Subtitles than with dual Chinese-English subtitles. Learners' self-assesed enjoyment of the viewing experience, as well as their comprehension of the video, both self-assessed and as indicated by independent evaluations of their summaries, remain unchanged.

Author Keywords
subtitles; interactive videos; language learning

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: Graphical User Interfaces

INTRODUCTION
Students studying foreign languages often wish to enjoy authentic foreign-language video content. For example, many students cite a desire to be able to watch anime in its original form as their motivation for starting to study Japanese. However, the standard presentations of videos 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 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 enjoyability of watching the content. Alternatively, the learner could simply watch a version with subtitles in their native language to enjoy the content. However, they would not learn the foreign language this way.

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:
• ACM copyright: ACM holds the copyright on the work. This is the historical approach.
• License: The author(s) retain copyright, but ACM receives an exclusive publication license.
• Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.
This text field is large enough to hold the appropriate release statement assuming it is single spaced.

We aim to build a foreign-language video viewing tool that maximizes vocabulary learning, while ensuring that the learner fully understands the video and enjoys watching it.

BACKGROUND
Videos in foreign languages have been adapted for foreign viewers and languages learners in many ways. These are summarized in Figure 1.

Presenting Videos to Foreign Viewers
One approach used to adapt videos to viewers who do not understand the original language is *dubbing*. Here, 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. Here, 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 little 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].

Presenting Videos to Language Learners
Whether or not videos should at all be presented to language learners with subtitles or similar aids is itself a matter of debate. Subtitles, in particular, are frowned upon, because they do nothing to deter learners from simply reading in their native language. Some language educators are of the opinion that since students will not have subtitles or similar aids when they visit the foreign country, then they should not be provided with any when viewing videos either. Nevertheless, various video comprehension aids have been experimented with in the context of language learning:

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 from conventional subtitles in the viewer's native language is that here, 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]

GliFlix is a variant on the conventional, native-language subtitle, which adds translations to the foreign language for the most common words that appear in the dialog. For example, for a French dialog, instead of "This is a line of dialog", GliFlix would show "This is a (un) line of dialog", showing that "a" in French is "un". In user studies with learners beginning to study French, they attain larger rates of vocabulary acquisition compared to regular subtitles, though not dual subtitles. Compared to dual subtitles, GliFlix has the disadvantage that because it shows only the most common vocabulary words in a dialog, then learners will not learn all the vocabulary in the video. Additionally, because GliFlix presents the foreign vocabulary in the order of the viewer's native language, this approach is likely less beneficial than dual subtitles for other language-learning tasks such as learning pronunciation and grammar.

SMART SUBTITLES 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.

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.

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.

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.

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.

IMPLEMENTATION

Smart Subtitles are automatically generated from captions with the assistance of dictionaries and machine translation. The Smart Subtitles system is implemented as 2 main parts: a system to extract subtitles and captions from videos, as well as an interactive interface to display them to learners.

Extracting Subtitles from Videos

Our system takes digital text captions in either the SubRip (SRT) or Web Video Text Tracks (WebVVT) formats as input. These are plain-text formats that specify a time range for each line of dialog, and the text that should be displayed. We can download these from various online services, such as

Universal Subtitles. However, many possible sources of subtitles either do not come with timing information, or are in non-textual formats, so we have developed a subtitle extraction system so that Smart Subtitles can be generated from a broader range of videos.

Extracting Subtitles from Untimed Transcripts

For many videos, a transcript is available, but the timing information stating when each line of dialog was said is unavailable. Examples include transcripts of lectures on sites such as OpenCourseWare, as well as lyrics for music videos.

It is possible to add timing information to videos automatically based on speech recognition techniques, which is called "forced alignment". However, we found that popular software for doing forced alignment yielded poor results on certain videos, particularly those with background noise and in foreign languages.

Thus, to generate timing information, we wrote an interface that plays the video, and the user is instructed to press the down button whenever a new line of dialog starts. We gather this data for several users to guard against user errors, and use it to compute the timing information for the transcript.

Extracting Subtitles from Overlayed-Bitmap Formats

Overlayed-bitmap subtitles are pre-rendered versions of the text which are overlayed onto the video when playing. They consist of an index mapping time-ranges to the bitmap image which should be overlayed on top of the video at that time. This is the standard subtitle format used in DVDs, where is it called VobSub, as well as other optical-disk video formats.

Because we cannot read text directly from the overlayed-bitmap images in DVDs, Smart Subtitles uses Optical Character Recognition (OCR) to extract the text out of each image. Then, it merges this with information about time ranges to convert them to the SRT subtitle format. Our implementation can use either the Microsoft OneNote [42] OCR engine, or the free Tesseract [44] OCR engine.

Extracting Subtitles from Hard-Subtitled Videos

Many videos come with hard subtitles, which include the subtitle as part of the video stream. Hard subtitles have the advantage that they can be displayed on any video player. However, hard subtitles have 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. Existing tools that perform this task, such as SubRip, are time-consuming, as they require the user to specify the color and location of each subtitle line in the video.

That said, hard subtitles are ubiquitous, particularly online. Chinese-language dramas on popular video-streaming sites such as Youku are almost always hard-subtitled in Chinese. Thus, to allow Smart Subtitles to be used with hard-subtitled videos, we devised an algorithm which can identify Chinese subtitles in hard-subtitled videos and extract them out.

Our hard-subtitle extraction algorithm takes advantage of properties of subtitles which we have found to 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.
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 many corners. This is also a Chinese-specific assumption, owing to the graphical complexity of Chinese characters.

Our hard-subtitle extraction first determines the color of the subtitle and its general location, extracts out pixels that are likely part of the subtitle, and applies a number of post-processing filters before running OCR on the extracted pixels to determine the subtitle text. Specifically, the algorithm is:

TODO do we actually want to describe the algorithm? It's pretty complex and is probably of little interest to CHI audiences...

The performance of the algorithm depends on the resolution of the video and the font of the subtitle. It generally works best on videos with 1280x720 or better resolution, and with subtitles that have distinct, thick outlines. The choice of OCR engine is also crucial - using Tesseract instead of OneNote more than tripled the character error rate, as Tesseract is much less resilient to extraneous pixels in the input.

Overall, on a set of 4 high-resolution Chinese hard-subtitled 5-minute video clips, the algorithm recognized roughly 80% of the dialog lines completely correctly, with roughly 95% of all characters correct. 2% of the errors at the dialog line level were due to the algorithm missing the presence of a line of dialog, as the OCR engine often failed to recognize text on lines consisting of only a single character or two. The remaining errors were due to character-level errors.

Unfortunately, because a single character-level error will often make the entire line of dialog incomprehensible to learners, in the Smart Subtitles application the dialog-level error rate is the metric of interest. The subtitles output by our algorithm thus require manual corrections before they can be presented to learners. However, the presence and timings of 98% of dialog lines were correctly recognized, so the information output by our algorithm helps with specifying subtitle timings.

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 dic-

tionary 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 [tabeta], though this word does not appear in the dictionary; only the infinitive form eat [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.

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].

Materials

The video we showed was the first 5 minutes, and the next 5 minutes, in the first episode of the drama (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 videos, which are filled with archaic vocabulary and expressions from literary Chinese. Additionally, the content of these video clips, consisting of conversations in classroom and household settings, 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.

Participants

Our study participants were 8 students who were enrolled in a third-semester Chinese class. 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. Participants were paid $20.

Research Questions

The questions our study seeked 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 viewers' self-assessed enjoyability differ between the tools?
4. Will viewers' self-assessed comprehension differ between the tools?
5. Will summaries viewers write about the clips after viewing differ in quality between the tools?
6. Which of the features of Smart Subtitles will users find helpful and actually end up using?

Procedure

Viewing Conditions

Half of the participants saw the first clip with dual subtitles and the second with Smart Subtitles, while the other half saw the first clip with Smart Subtitles and the second with dual subtitles. For the dual subtitles condition we used the KM-Player video player, showing English subtitles on top and Chinese on the bottom. For the Smart Subtitles condition we used our software.

Before participants started watching each clip, we informed them that they would be given a vocabulary quiz afterwards, and that they should attempt to learn vocabulary in the clip while watching the video. We also showed them how to use the video viewing tool during a minute-long familiarization session on a seperate clip before the session. Participants were told they could watch the clip for as long as they needed, pausing and rewinding as they desired.

Vocabulary Quiz

After a participant finished watching a clip, we evaluated vocabulary learning 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 foreign-language learning [12].

Questionnaire

After completing the vocabulary quiz, we asked them to write a summary of the clip they had just seen, describing as many details as they could recall. Then, they completed a questionnaire where they rated on a 7-point likehardt scale, the following questions:

- 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?

Finally, we asked for free-form feedback about the user's impressions of the tool, and whether they would use the tool themselves.

RESULTS

From our study, we found that:

1. Users learned over twice as much vocabulary using Smart Subtitles than with dual subtitles?
2. Viewing times did not differ significantly between tools.
3. Viewers' self-assessed enjoyability did not differ significantly between tools.
4. Viewers' self-assessed comprehension did not differ significantly between the tools.
5. **TODO Will summaries viewers write about the clips after viewing differ in quality between the tools?**
6. Users made extensive use of both the word-level translations and the dialog-navigation features of Smart Subtitles, and described these as helpful.

Vocabulary Learning

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.

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. Interestingly, 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.

Self-Assessment Results

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). Free-form feedback from participants indicates that an increased perceived ability to follow the original Chinese dialog contributed to the enjoyability result.

Summary Quality Ratings

After watching each video, partcipants wrote a summary describing the clip they had seen. An example is:

It was about a failed teacher whose students don't take him seriously (they leave his class, want to beat him up), and then a president who is angry about his daughter doing poorly at math after hiring an expensive tutor.

To evaluate the quality of the summaries written by our participants, we hired 5 Chinese-English bilingual raters to rate the summaries. The raters were hired from the oDesk contracting site, and were paid $15 apiece. Raters were first asked to view the clips, and write a summary in English to show that they had viewed and understood the clips. Then, we presented them the summaries written by students in random order. For each summary, we indicated which clip was being summarized, but the raters were blind as to which condition the student had viewed the clip under. Raters were asked to rate, on a scale of 1 (worst) to 7 (best):

- From reading the summary written by the student, how much does the student seem to understand this part of the clip overall?
- How many of the major points of this part of the clip does their summary cover?
- How correct are the details in this summary of this part of the clip?
- How good a summary of this part of the clip do you consider this to be overall?

TODO statistical significance and inter-rater agreement stuff

Free-form Feedback

We asked users to provide feedback about the watching experience in general, and whether they were interested in using the tool again. Of our 8 users, all expressed interest in using Smart Subtitles again. The written feedback that participants wrote indicated that they found most of the interface features to be helpful. 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 pinyin, 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.

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.

Viewing strategies with Smart Subtitles varied across participants, though all made at least some use of both the word-level and phrase-level translation functionality. Word-level translations were heavily used. On average, users hovered over words in 3/4 of the lines of dialog (standard deviation 0.22). The words hovered over the longest tended to be less common words, indicating that participants were using the feature for defining unfamiliar words, as intended. Participants tended to use phrase-level translations sparingly. On average they clicked on the translate button on only 1/3 of the lines of dialog (standard deviation 0.15). Combined with our observation that there was no decline in comprehension levels with Smart Subtitles, this suggests that word-level translations are often sufficient for learners to understand dialogs.

CONCLUSION AND FUTURE WORK

We have presented Smart Subtitles, an interactive transcript with features to help learners, such as vocabulary definitions on hover and dialog-based video navigation. They can be automatically generated from common sources of videos and subtitles, such as DVDs.

Our user study found that participants learned more vocabulary with Smart Subtitles than dual Chinese- English subtitles, and rated their comprehension and enjoyment of the video as similarly high. Independent ratings of summaries written by participants further confirm that comprehension levels when using Smart Subtitles match those when using dual subtitles. Given that users only viewed phrase-level translations for a third of the dialog lines when using Smart Subtitles, yet matched their comprehension levels with dual Chinese-English subtitles, this suggests that word-level translations are often sufficient for comprehension.

Unlike traditional subtitles, Smart Subtitles currently expect users to actively interact with them. However, we could potentially allow more passive usage by using statistical modelling to predict which words the viewer wont know and automatically showing their definitions.

Much work can still be done in the area of incorporating multimedia into learning. Our current Smart Subtitles system focuses on written vocabulary learning while watching of dramas and movies. However, we believe that augmenting video can also benefit other aspects of language learning. For example, we could incorporates visualizations for helping learn grammar and sentence patterns, and speech synthesis for helping learn pronunciation. We could also pursue further gains in vocabulary learning and comprehension, by dynamically altering the video playback rate, or by adding quizzes into the video to ensure that the user is continuing to pay attention.

Other multimedia forms can likewise benefit from interfaces geared towards language learning, though each form comes with its own unique challenges. For example, the current Smart Subtitles system can easily be used with existing music videos and song lyrics. However, the system would be even more practical for music if we could remove the need for an interactive display, and simply allowed the user to learn while listening to the music. Multimedia that is naturally interactive, such as Karaoke, likewise presents interesting opportunities for making boring tasks, such as practicing prononciation, more interesting to learners.

We hope our work leads 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.

in particular, pronunciation, listening comprehension ability, and grammar learning.

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

Such text-format captions can be found for some videos on sites such as Many sources of subtitles and captions, however

The Smart Subtitles system is able to use subtitles and captions in several formats.

On each page your material (not including the page number) should fit within a rectangle of 18 x 23.5 cm (7 x 9.25 in.), centered on a US letter page, beginning 1.9 cm (.75 in.) from the top of the page, with a .85 cm (.33 in.) space between two 8.4 cm (3.3 in.) columns. Right margins should be justified, not ragged. Beware, especially when using this template on a Macintosh, Word can change these dimensions in unexpected ways. Please be sure that your PDF is US letter and not A4. If your PDF or paper are formatted for A4, the submission will be returned to you to fix.

TYPESET TEXT

Prepare your submissions on a word processor or typesetter. Please note that page layout may change slightly depending upon the printer you have specified. LaTeX sometimes will create overfull lines that extend into columns. To attempt to combat this, the .cls file has a command, \sloppy, that essentially asks LaTeX to prefer underfull lines with extra whitespace. For more details on this, and info on how to control it more finely, check out **http://www.economics.utoronto.ca/osborne/latex/PMAKEUP.HTM**.

Title and Authors

Your paper's title, authors and affiliations should run across the full width of the page in a single column 17.8 cm (7 in.) wide. The title should be in Helvetica 18-point bold; use Arial if Helvetica is not available. Authors' names should be in Times Roman 12-point bold, and affiliations in Times Roman 12-point. For more than three authors, you may have to place some address information in a footnote, or in a named section at the end of your paper. Please use full international addresses and telephone dialing prefixes. Leave one 10-pt line of white space below the last line of affiliations.

Abstract and Keywords

Every submission should begin with an abstract of about 150 words, followed by a set of keywords. The abstract and keywords should be placed in the left column of the first page under the left half of the title. The abstract should be a concise statement of the problem, approach and conclusions of the work described. It should clearly state the paper's contribution to the field of HCI.

The first set of keywords will be used to index the paper in the proceedings. The second set are used to catalogue the paper in the ACM Digital Library. The latter are entries from the ACM Classification System [**?**]. In general, it should only be necessary to pick one or more of the H5 subcategories, see **http://www.acm.org/class/1998/ccs98.html**

Normal or Body Text

Please use a 10-point Times Roman font or, if this is unavailable, another proportional font with serifs, as close as possible in appearance to Times Roman 10-point. The Press 10-point font available to users of Script is a good substitute for Times Roman. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times and not Times New Roman. Please use sans-serif or non-proportional fonts only for special purposes, such as headings or source code text.

First Page Copyright Notice

Leave 3 cm (1.25 in.) of blank space for the copyright notice at the bottom of the left column of the first page. In this template a floating text box will automatically generate the required space. Note however that the text box is anchored to the **ABSTRACT** heading, so if that heading is deleted the text box will disappear as well. You can replace the default copyright notice by uncommenting the \toappear block at the beginning of the document and inserting your own text, for example, for versions under review.

Subsequent Pages

On pages beyond the first, start at the top of the page and continue in double-column format. The two columns on the last page should be of equal length.

Figure 1. With Caption Below, be sure to have a good resolution image (see item D within the preparation instructions).

References and Citations

Use a numbered list of references at the end of the article, ordered alphabetically by first author, and referenced by numbers in brackets [**?**, **?**, **?**, **?**]. For papers from conference proceedings, include the title of the paper and an abbreviated name of the conference (e.g., for Interact 2003 proceedings, use *Proc. Interact 2003*). Do not include the location of the conference or the exact date; do include the page numbers if available. See the examples of citations at the end of this document. Within this template file, use the `References` style for the text of your citation.

Your references should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible (i.e., you provide the address for obtaining the report within your citation) and may be obtained by any reader for a nominal fee. Proprietary information may not be cited. Private communications should be acknowledged in the main text, not referenced (e.g., "[Robertson, personal communication]").

Objects	Caption — pre-2002	Caption — 2003 and afterwards
Tables	Above	Below
Figures	Below	Below

Table 1. Table captions should be placed below the table.

SECTIONS

The heading of a section should be in Helvetica 9-point bold, all in capitals. Use Arial if Helvetica is not available. Sections should not be numbered.

Subsections

Headings of subsections should be in Helvetica 9-point bold with initial letters capitalized. For sub-sections and sub-subsections, a word like *the* or *of* is not capitalized unless it is the first word of the heading.)

Sub-subsections

Headings for sub-subsections should be in Helvetica 9-point italic with initial letters capitalized. Standard \section, \subsection, and \subsubsection commands will work fine.

FIGURES/CAPTIONS

Place figures and tables at the top or bottom of the appropriate column or columns, on the same page as the relevant text (see Figure 1). A figure or table may extend across both columns to a maximum width of 17.78 cm (7 in.).

Captions should be Times New Roman 9-point bold. They should be numbered (e.g., "Table 1" or "Figure **??**"), centered and placed beneath the figure or table. Please note that the words "Figure" and "Table" should be spelled out (e.g., "Figure" rather than "Fig.") wherever they occur.

Papers and notes may use color figures, which are included in the page limit; the figures must be usable when printed in black and white in the proceedings. The paper may be accompanied by a short video figure up to five minutes in length. However, the paper should stand on its own without the video figure, as the video may not be available to everyone who reads the paper.

LANGUAGE, STYLE AND CONTENT

The written and spoken language of SIGCHI is English. Spelling and punctuation may use any dialect of English (e.g., British, Canadian, US, etc.) provided this is done consistently. Hyphenation is optional. To ensure suitability for an international audience, please pay attention to the following:

- Write in a straightforward style.
- Try to avoid long or complex sentence structures.
- Briefly define or explain all technical terms that may be unfamiliar to readers.
- Explain all acronyms the first time they are used in your text—e.g., "Digital Signal Processing (DSP)".
- Explain local references (e.g., not everyone knows all city names in a particular country).
- Explain "insider" comments. Ensure that your whole audience understands any reference whose meaning you do not describe (e.g., do not assume that everyone has used a Macintosh or a particular application).
- Explain colloquial language and puns. Understanding phrases like "red herring" may require a local knowledge of English. Humor and irony are difficult to translate.
- Use unambiguous forms for culturally localized concepts, such as times, dates, currencies and numbers (e.g., "1-5-97" or "5/1/97" may mean 5 January or 1 May, and "seven o'clock" may mean 7:00 am or 19:00). For currencies, indicate equivalences—e.g., "Participants were paid 10,000 lire, or roughly $5."
- Be careful with the use of gender-specific pronouns (he, she) and other gendered words (chairman, manpower, man-months). Use inclusive language that is gender-neutral (e.g., she or he, they, s/he, chair, staff, staff-hours, person-years). See [**?**] for further advice and examples regarding gender and other personal attributes.
- If possible, use the full (extended) alphabetic character set for names of persons, institutions, and places (e.g., Grønbæk, Lafreniére, Sánchez, Universität, Weißenbach, Züllighoven, Århus, etc.). These characters are already included in most versions of Times, Helvetica, and Arial fonts.

ACCESSIBILITY

The Executive Council of SIGCHI has committed to making SIGCHI conferences more inclusive for researchers, practitioners, and educators with disabilities. As a part of this goal, the all authors are asked to work on improving the accessibility of their submissions. Specifically, we encourage authors to carry out the following five steps:

1. Add alternative text to all figures
2. Mark table headings
3. Add tags to the PDF
4. Verify the default language
5. Set the tab order to "Use Document Structure"

Unfortunately good tools do not yet exist to create tagged PDF files from Latex. LaTeX users will need to carry out all of the above steps in the PDF directly using Adobe Acrobat, after the PDF has been generated.

For more information and links to instructions and resources, please see: **http://chi2014.acm.org/authors/guide-to-an-accessible-submission**.

PAGE NUMBERING, HEADERS AND FOOTERS

Please submit your anonymous version for reviewing with page numbers centered in the footer. These must be removed in the final version of accepted papers, as page numbers, headers, and footers will be added by the conference printers. Comment out the \pagenumbering command at the top of the document to remove page numbers.

PRODUCING AND TESTING PDF FILES

We recommend that you produce a PDF version of your submission well before the final deadline. Your PDF file must be ACM DL Compliant. The requirements for an ACM Compliant PDF are available at: `http://www.sheridanprinting.com/typedept/ACM-distilling-settings.htm`.

Test your PDF file by viewing or printing it with the same software we will use when we receive it, Adobe Acrobat Reader Version 7. This is widely available at no cost from [**?**]. Note that most reviewers will use a North American/European version of Acrobat reader, which cannot handle documents containing non-North American or non-European fonts (e.g. Asian fonts). Please therefore do not use Asian fonts, and verify this by testing with a North American/European Acrobat reader (obtainable as above). Something as minor as including a space or punctuation character in a two-byte font can render a file unreadable.

BLIND REVIEW

For archival submissions, CHI requires a "blind review." To prepare your submission for blind review, remove author and institutional identities in the title and header areas of the paper. You may also need to remove part or all of the Acknowledgments text. Further suppression of identity in the body of the paper and references is left to the authors' discretion. For more details, see the submission guidelines and checklist for your submission category.

CONCLUSION

It is important that you write for the SIGCHI audience. Please read previous years' Proceedings to understand the writing style and conventions that successful authors have used. It is particularly important that you state clearly what you have done, not merely what you plan to do, and explain how your work is different from previously published work, i.e., what is the unique contribution that your work makes to the field? Please consider what the reader will learn from your submission, and how they will find your work useful. If you write with these questions in mind, your work is more likely to be successful, both in being accepted into the Conference, and in influencing the work of our field.

ACKNOWLEDGMENTS

We thank CHI, PDC and CSCW volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document. Some of the references cited in this paper are included for illustrative purposes only. **Don't forget to acknowledge funding sources as well**, so you don't wind up having to correct it later.

REFERENCES FORMAT

References must be the same font size as other body text.