Subtitle-based Interactive Exercises

Thomas Plagwitz

Subtitle-based Interactive Exercises 1
Thomas Plagwitz 1
1 Introduction 2
1.1 Pedagogical Need 2
2 Scope: still focussing on automation = relief of lower-level tasks 3
3 “How the verbal exercises work in tandem with the film itself” = More of a life cycle of a weekly class 4
3.1 Template-based exercises for self-access/homework 5
3.2 Face to face class 8
3.2.1 Quiz.dot template, including lookup for lower (e.g. vocabulary) language skills 8
3.2.2 PowerPoint Template for higher level (comprehension) language skills 9
3.2.3 Timeline Template for higher level (collaborative and comparative) plot-summary 9
3.3 Business Case 5
4 Procedure 5
5 “technical limitations “, relieving the teachers from “the user-friendly (or unfriendly) interface between the material and the design” or helping them “get used to it” 10
5.1 Back office 10
5.2 Teacher training 11
6 Outlook 11
7 Appendix 14
8 References 24

1 Introduction

1.1 Pedagogical Need

In our era of “Iconic turn”, English teaching programs tend to own and use a large number of DVDs as well have access to other – increasingly digital[1] - archived off-the-air video recordings which the UK copyright stipulations for educational institutions allows to liberally reuse during instruction, which are highly popular with the language learning student audience. While audio-visual materials add valuable channels to instruction, for true multimedia, the text channel should not be neglected either, to counterbalance the current "Verlagerung von der sprachlichen auf die visuelle Information, vom Wort auf das Bild und – am beunruhigendsten – vom Argument auf das Video"[2].Per se, these source materials consumer market DVDs and video transmission were not designed for foreign language learners. E.g. they seldom come with target language subtitles; even if they do, viewing target language subtitles in addition to listening to the audio does not provide sufficient aid in understanding and learning for most proficiency levels.

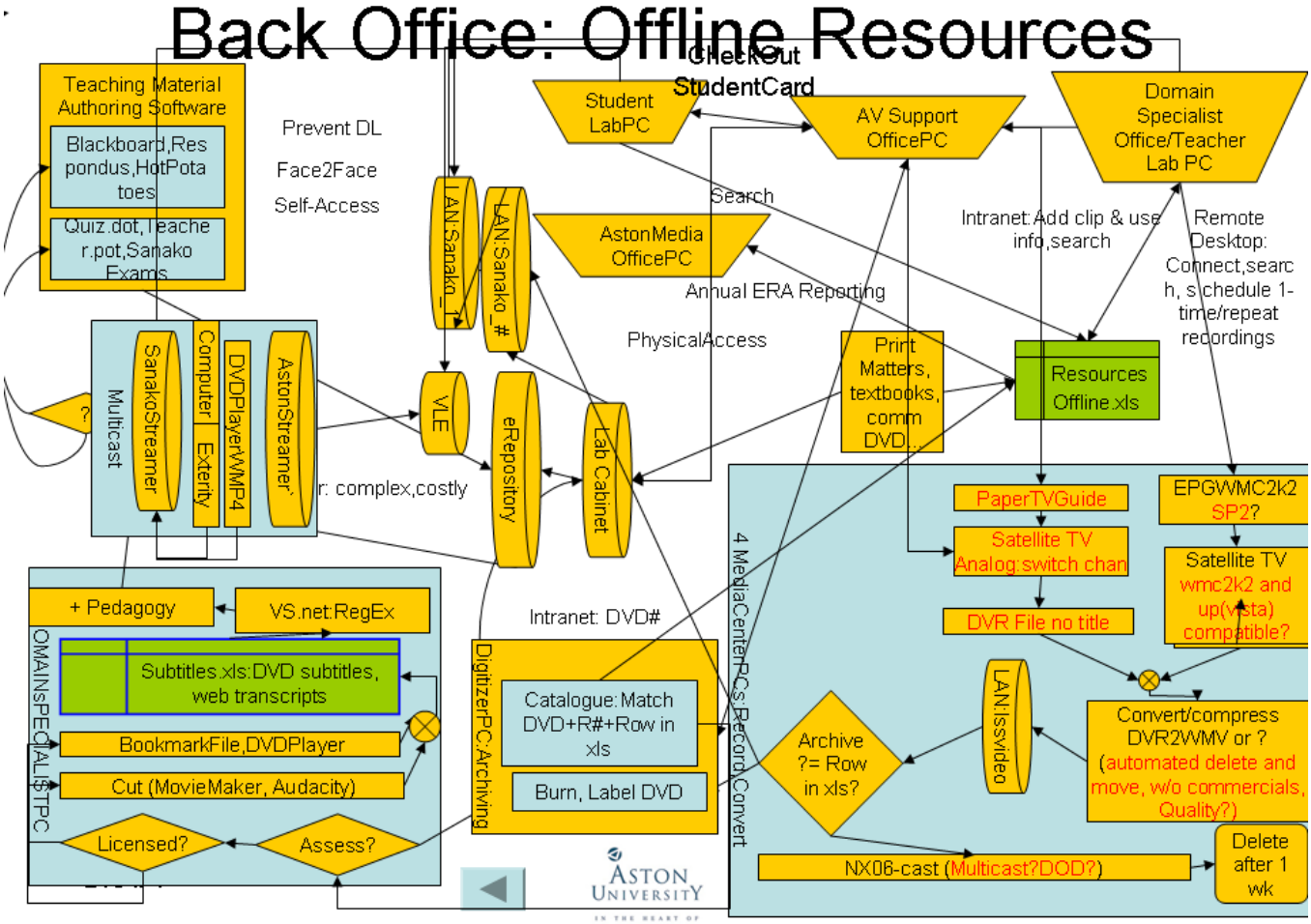
Showing L1 translation subtitles is a last resort of limited pedagogical value. Hardly any other current pedagogical (textbook) materials that are used in foreign language instruction resort to translations after the first few weeks of language instruction.

On the other hand, glossing with L2 information, is a common pedagogical tool in aiding comprehension. In previous posts, I created Blackboard exercises for two German DVD movies that split up these movies into 175 scenes and essay questions and included the complete transcripts (based on the subtitles) and a model plot summary (based on a scene protocol).

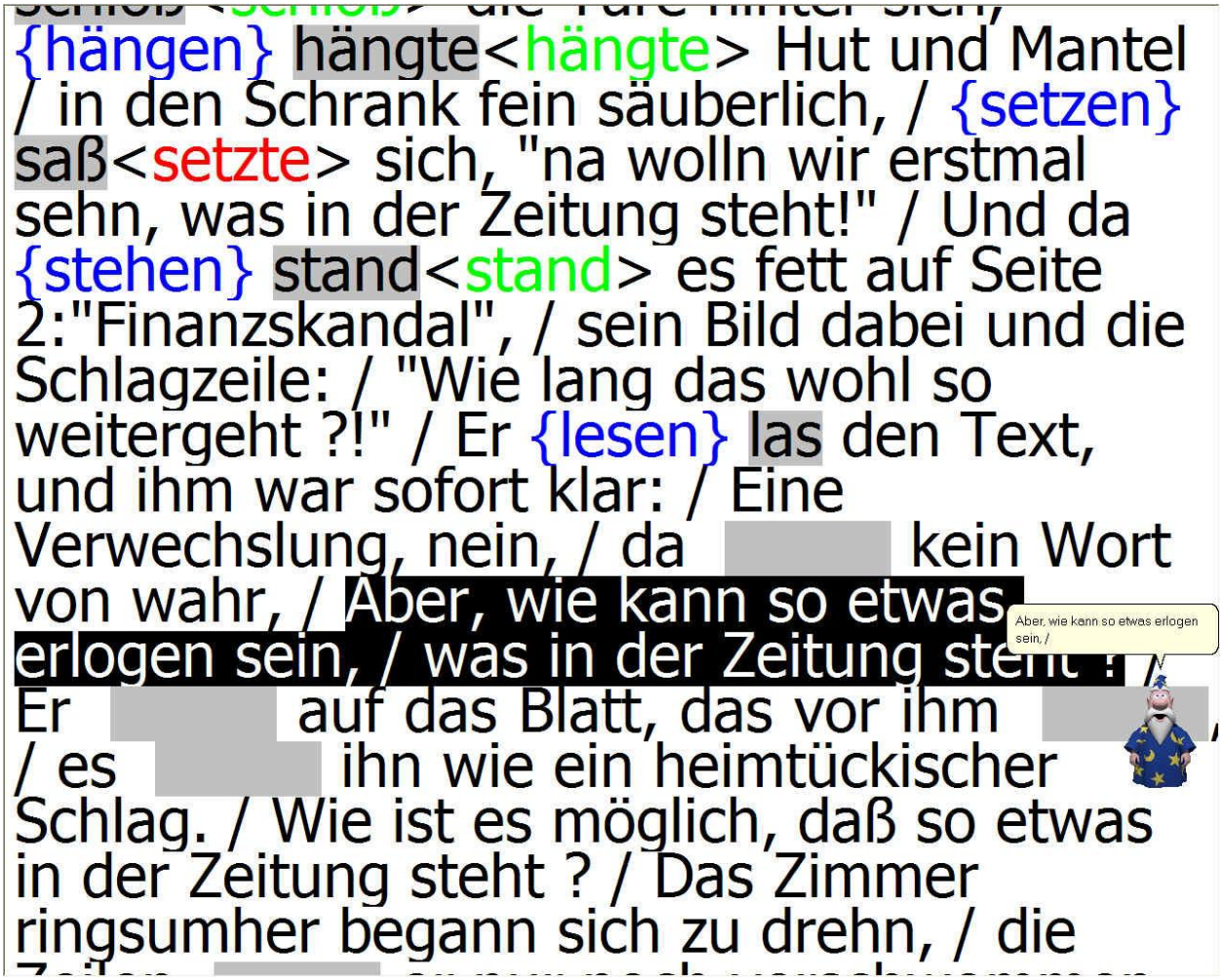
I would like to propose for creating MS-Word based exercises for 6-8 more movies for the major languages in the departmental program, as exemplified in the pictures below.

---

[1] The opportunity which UK copyright law, and advances in media availability (English and international free-to-air digital Satellite TV channels with metadata) and computer technology in hardware (storage media capacity) and software (Windows Media Center) may not have been realized in many institutions. I outline their potential for recording and creation of teaching materials in this flowchart:



[2] Sauerländer, Iconic turn? Eine Bitte um Ikonoklasmus. Vortrag im Rahmen der Vorlesungsreihe "Iconic Turn" an der Ludwig-Maximilians-Universität München. 10. Juli 2003. URL: http://netzspannung.org/cat/servlet/CatServlet?cmd=netzkollektor&subCommand=showEntry&entryId=104989&lang=de.



2 Scope: still focusing on automation = relief of lower-level tasks

I wish to clarify am not trying to propagate the use of feature films in ELT teaching which seems pretty well-established. Nor should there be much need to call attention to the availability of pre-digitized text of professional writer quality and considerable cultural significance, in the form of movie subtitles. I am also not denying – and hope to some extend understand, and partially have taught using such exercise activities in the higher, more complex aspects of language mastery, for years, and, albeit to a much lesser extend, also designed such exercises.

However, what I am trying to focus here on (in other words: specialize in, or in other words again: how I try to give evidence of my modesty) is the application of computer technology (in the form of automation) to ELT exercise creation. I consider this application as much "die Forderung des Tages"[3] to knowledge worker workflow processes in the ongoing "mechanization of intelligence" as the application of steam to production processes during the "mechanization of physical labor" in the Britain of the late 18th and early 19th century. Which implies not pouring steam over fabric to see it shrink and shrivel and then proclaim "Well, see? That can't work!", but funneling it through a rather complex (we have oodles lots of time as a species if only we learn to collaborate and reuse) piping system to drive simple movements which result in (initially even coarser) output materials, but are cheap and free the formerly manual toilers to do things which steam engines cannot (yet) do – including "conversational gambits etc. , pragmatic features of spoken communication".

It seems beneficial that ESL departments use such outsourced automated services, may be even include a specialist who funnels the various forms of steam from such services into such Given the fact that in reality ESL teachers still have to deal with more basic language skills and still have to invest time in manual creation of exercises for such basic language skills (or even worse, at least in the reality I have experienced in SLA institutions,. rather waste time fighting over not to have to this), the automation of such exercise creation seems to be beneficial.

The classical example is the prevalence of the 2nd generation Google search technology over the more short-circuited thinking of Yahoo catalogue paradigm: human toiling is superseded by automation – which uses cheap - especially considering the ROI -, incredibly cheap CPU cycles - even if they may initially only involve some *prima facie* rather lowly, infrastructural operations like retrieving, sorting (number crunching!), searching (string matching!), filtering -, or rather of automated harvesting and reuse/recycling of human intelligence. While there is more to the Google page rank algorithm than

[3] "Was aber ist deine Pflicht? Die Forderung des Tages" (Johann Wolfgang Goethe, Maximen und Reflexionen. Allg., Ethisches, Lit. II, 31. Quoted in http://www.wissen-im-netz.info/literatur/goethe/maximen/1-02.htm.

the string-matching and counting of a basic “retrieve all links on the WWW, what gets manually (by human intelligence) linked much, gets ranked first in searches (especially if search terms appear in the link title and in the link context and …)” (not at least the 2nd application of this principle: automated gathering and digesting of the feedback how human intelligence interacts with these search results by clicking through, and since more recently also by explicit “comment/promote/remove”), the market outcome seems to be evidence of this combination of human and artificial intelligence being a very powerful approach.

3 “How the verbal exercises work in tandem with the film itself” = More of a life cycle of a weekly class

It is a basic rule of PRINCE2 project management methodology to include in an initial project feasibility assessment and cost/benefit analysis the radical option: “What if you do nothing?”. So: What was the SLA situation without these template-based exercises I developed?

a SLA department head who let me know when I started there “They do not know how we teach”, had led a fallout over “Can we haz TV on each students’ lab computer?”[4] (we could not, for cost/benefit reasons). Consequently, she could be observed in the language lab, using a VCR and the screen projector,[5] while ignoring the networked computers and language lab software,[6] letting her classes watch target language feature films with a handful of comprehension questions for two class hours while correcting student assignments in the rear of the classroom.

Since this is an above average example of the reality of SLA departments I worked in, it seems best to take “people out of the equation”[7] and have as much automation as possible in the SLA instruction and as little need for the teacher to produce anything. Possible ways to use the template-based exercises to improve her classes would be these:

This is where your work could really offer something new and very useful in the development of film in language learning. Of course, in so far as you have provided materials for using your software, you have designed something useful with regard to teaching materials, but the rationale is still too oriented towards a technical readership and the exercises you have developed are too micro and too located in the written text that it almost feels as if the film itself is no more than a carrier of grammatical structures!! 40%

---

[4] It may serve as evidence for the much improved feasibility of a subtitle-based teaching material creation that it took me only a minute or so to locate and download subtitles in a variety of formats, a well as for a variety of versions this film has been distributed in, as well as for the effectiveness of using film subtitles for ELT that I still remember that “[t]he broadcast of the '36 Olympics was the first TV transmission.....that went into space”: “"I declare the Games in Berlin.....at the celebration.....of the first Olympics of the New Era.....as opened." Despite the popularity of inundating students with TV broadcasts not only in SLA pedagogy, but in lecture recording in general, I would like to challenge the belief that we need more of that Nazi technology in SLA. I created my subtitle template-based exercises to, in practice, enable some interactivity opportunity in this “Es spricht: Der Führer!” pedagogy which students are routinely are confronted withI know but too well from experience how convenient this is for the teacher, when I had no time to prepare class due to having to plumb infrastructure, I just talked in foreign tongues), as well in practice (maybe) have a better quality intervention, primarily more meaningful, of the teacher in student learning, if only she raises her eye from correcting student papers in class when my templates flash red and she finds an error worth commenting on and we can teach her how easy it is to connect to a student for a quality time intervention. In other words, I hold little hope of, by showing her ideal models, getting (not only) this department head to better prepare and deliver her SLA classes. Rather, in my experience – which may go contrary to common belief - it is much easier to make computers produce new things, like these (mostly) automated exercises, than SLA teachers (this may have partially to do with computers being heavily underused and SLA teachers not being willing to be cajoled anymore in their rut. Am I/is technology lending a hand to administrators that are eager to cut SLA resources, make SLA instruction get by on the cheap? In my experience, productivity gains get divided up, one part for administrators, one part for teachers, some for the students’ benefit – or so I hope, but I do not feel I need to help SLA teachers negotiate their share).

[5] Strinkingly, the big screen was also more popular with the Nazis than TV.

[6] I do not remember whether she displayed the audio over the language lab headsets. I do remember she wanted a classroom speaker installed (despite it tending to cause feedback in language lab system) and using the TV set loudspeakers a replacement. I do not think she preferred the students listening over the headphones even while she was correcting assignments, she just sat in the rear.

[7] “My big belief in management is that people don’t change. You know, I’m highly unlikely at this age to fundamentally change what I am as a human being, and so my management philosophy also tends to be that if I can complement people’s strengths by surrounding them with people who can complement their areas of weakness, that’s probably a better recipe for success than trying to say, ‘‘O.K., you need to change.’” (Shantanu Narayen, president and chief executive of Adobe Systems, quoted after http://www.nytimes.com/2009/07/19/business/19corner.html?_r=1&ref=technology). This senior management insight is only slightly different from the more general “boys will be boys”, “SLA teachers will be SLA teachers”. Since “surrounding people” we do not have, as they are expensive (much more than machines), we better try and find a use for the latter.

3.1 Lookup Template-based exercises for self-access/

3.1.1 Business Case

- The templates can improve the return on the following usually **existing investments**.
  - Language Lab
    - Library of DVDs and off-air recordings
      - I do not anticipate commercial publishers providing ancillary materials for their DVDs that would make the materials which will be produced in this project redundant.
    - Student computers
      - Students can take advantage of the lab computers by getting their individual vocabulary questions answered when watching the movie.
      - Digital Audio Lab Systems allow for "transferring" the video audio to the class group, while having the class group "group-conference" and "mixing" in the teacher voice. This can provide for "better than nature" exposure to spoken target language in classes revolving around these exercises and constitute an intensive and well rounded use of a language lab environment.
  - The project is designed to impact the maximum **number of students and courses.** Beginners and intermediate student can practice: Key words or new vocabulary or words that illustrate specific grammar points

3.1.2 Procedure

- Listening comprehension exercises
  - Is the default use. You can quiz for any terms students have to comprehend form “dictation”.
  - You can however give there exercise a topical twist by omitting certain types of words. You can automate this by building (and archiving for reuse) regular expressions and running a replacement operation against the subtitle file of choice.
  - Simpler regular expressions can build grammar exercises focusing on function words; more complicated regular expressions would be needed for building vocabulary exercises focusing on semantic fields
  - Sets of words which can be handled relatively easily by regular expressions,[8] are so-called closed class words.[9]
    - Grammar exercises:
  - On the other hand, WordNet – a resource which is being increasingly used for automated generation of ESL SLA exercises generation [10]- “only contains ‘open-class words’: nouns, verbs, adjectives, and adverbs. Thus, excluded words include determiners, prepositions, pronouns, conjunctions, and particles.”[11]
    - “WordNet is organized by the concept of synonym sets (synsets), groups of words that are roughly synonymous in a given context. The glossary definition and the example sentences are shared among all synonyms in a given synset.”[12]

---

[8] A regular expression can describe a group of different matches with one string, containing wildcards which are smaller strings that can be matched by 0-many characters..

[9] “In linguistics, a **closed class** (or **closed word class**) is a word class to which no new items can normally be added, and that usually contains a relatively small number of items. Typical closed classes found in many languages are adpositions (prepositions and postpositions), determiners, conjunctions, and pronouns.” (http://en.wikipedia.org/wiki/Closed_class). Also see http://en.wikipedia.org/wiki/Function_word: “Function words belong to the closed class of words in grammar in that it is very uncommon to have new function words created in the course of speech”; “Function words (or grammatical words) are words that have little lexical meaning or have ambiguous meaning, but instead serve to express grammatical relationships with other words within a sentence, or specify the attitude or mood of the speaker.”.

[10] TBA mailing list – See also http://wordnet.princeton.edu/links#extensions : “Tony Veale of University College Dublin is heading KNOW-BEST, which is attempting to use WordNet and other lexical ontologies as the generative basis for on-line language games. Playable samples of the games and information/papers related to their construction/research issues are available at The Creative Language Systems Group web site.”

[11] http://wordnet.princeton.edu/faq#cat_missing.

[12] http://wordnet.princeton.edu/faq#cat_content.

        - Vocabulary exercises: If you can derive from that topical classes and apply a word frequency filter, you can probably generate regular expressions, incuding for inflected forms, that can generate exercises for the vocabulary introduced in typically topically organized chapters in beginner and intermediate learner textbooks (or possibly textbooks should bas3e their organization on the wordnet results).

- Convert.sln: An overview of this program (under development) is in Figure
    - The Principle of the program: The program tries to provide a framework for text manipulation (replacement)to which specific replacement operations that are deemed useful in SLA (and management of an e-learning service facility for SLA) can be added over time
    - Respondus:
        - the origin of the project was an attempt to convert raw text into the format required for the ms-word files on the basis of which Respondus® addin automates the tedious (click and wait) generation of assessments in Blackboard®
        - in this case semi-automatically generated segment in an Excel –file, which split up a lengthy cultural history documentary series in a series of scenes, each of which the student viewers have to answer the journalistic questions for, to document viewing comprehension, and then compare their input to the prior description of the scene by their instructor – no automatic evaluation possible for this type of question, but an instructive model for the student, and simplified manual corrections for the instructor.[13]
    - Subtitles: Prepare the subtitle source file
        - the input format of records in the SRT, one of the most common subtitle formats,
            - has a certain evolved way to indicate the times when subtitles start and stop being visible on screen
            - mimics, with line breaks, the way the subtitles are displayed on screen
        - the regular expressions
            - first divide the subtitle into matches of one “record” each
            - using delegate programming, post-processes each individual record for more advanced replacements (line break removal)
        - the output format:
            - An example, based on subtitles for “Miami Vice” , is shown in Figure 2: Windiff of MassageSubtitlesSRT where red is the original format (end of input section) and yellow is the output format (beginning of section, so no direct equivalent of the content).
            - simplifies the input format
            - and allows for
                - with minor modifications (replacing f the asterisks within the html-brackets, and surrounding the entire table with <table> tags and a valid HTML <head><body> frame), display as a HTML table (see Figure 6: Subtitles Reformatted as HTML Table)
                - and consequently easy importing into e.g. excel (see Figure 3: Subtitles converted to XLS)
                - and easy post-editing in there (see included file centre_overall\quiz_dot_create_excel.wmv),
                    - e.g. removal of stop times,
                    - simplification of timestamps
                    - filtering on time, so that a subtitle file can be easily divided into e.g. 5-minute viewing installments
    - Function words: A number of reusable regular expressions for typical learner difficulties in a variety of languages other than English have already been produced:

[13] See Figure 4: Glossed Subtitle VLE Assessment, Student answering and Figure 5: Glossed Subtitle VLE Assessment, Instructor model answer.

- German
    - Produces close quizzes where students have to decide whether a personal pronoun in the accusative or dative case has to be filled in
- Spanish
    - Produces close quizzes where students have to decide which form of the demonstrative pronoun should be filled in
- Italian
    - Produces close quizzes where students have to decide whether a preposition and article or contraction of both should be filled in
- I now added an English one regarding articles
    - Choice
        - One of of the most frequent errors in English by learners from L1 without an article system (East Asian and Slavic languages) is the “a or an/the” dichotomy.[14]
        - Fortunately, this is also a very simple replacement (admittedly hardly requiring a regular expression).
        - Regex:
            - Match: (?:boundary1\b)(?:article(a|A)n*|(t|T)he)(?:boundary2\b)
                - Post-processing by the VBA string comparison in quiz.dot will be not case sensitive, but we still want to match case-sensitive matches here since we do not want to match possible ACRONYM (e.g. “THE” = “Times Higher Education”)
            - Replacement: \<Definite or indefinite article?\>\{\boundary1\article\boundary2\}
            - Results: gap should be produced for
                - 160 forms of the definite article
                - 107+7 forms of the indefinite article
    - Quiz.dot
        - The results get pasted into the template quiz.dot
        - The macro generatequiz is run against the template (see Figure 7: subtitles ready for quiz generation in quiz.dot)
        - Result (included): **MIAMI_VICE_21_text.doc** (can be run on any computer with MS-Office and Macro Security set to “medium”and macros enabled on startup).
    - Usage:
        - the quiz can be used as a listening comprehension quiz while watching the movie, reading along and tabbing from article to article gap (see Figure 8: Quiz as viewed by Student and Figure 9: Quiz result summary, as viewed by student)[15]
        - The movie can also be viewed
            - in short segments with an instructor using the powerpoint to
                - Introduce, explain beforehand vocabulary, which the students then have to locate in its context
                - Ask comprehension questions
                - Possibly both (watch each segment twice)

[14] See (Muto-Humphrey), 61. The author also finds that learnes find it difficult to apply the rules governing the English article system and that it is therefore ““necessary to learn the articles under the situation where they are actually used by native speakers” which is what my article exercise based on movies in English would do.

[15] There are last gaps here than expected, as I MS-Word ran into an out of memory condition. Normailly, I have not used more than 50 gaps pe file, as the quiz tends to get slow. This does not seem to be the case here, but I would for final production still split up the input subtitles into segments.

- And the in a review, the students fill out the quiz, where the focus is not on mere listening comprehension, but on being able to memorize the model and to reproduce it correctly.

3.1.3 Lookup integration

- The templates include a dictionary/encyclopedia lookup feature, which can be either accessed by
  - selecting the word with the mouse and select the lookup source of choice form a MS-Word menu
  - double-clicking the word in MS-Word which looks it up in the default online dictionary for the language the word is set to in MS-Word.
  - This feature automates the glossing which SLA teachers (are supposed to) do when not using pre-built materials (of which there are not enough).
    - This workflow should be improved by keeping track (= a personal log file) of the students' lookups (and possibly

      creating review exercises on the basis of words, or better: the context the student has encountered them in the subtitle text for the student

      a way for the teacher to access this log (ideally cumulative) to (formatively) assess the students' vocabulary acquisition and adapt additional, less automated teaching/exercises to address remaining issues.

3.2 Face to face class

3.2.1 Quiz.dot template, including lookup for lower (e.g. vocabulary) language skills

The lower-level language skill exercises can help relieve her the teacher from some of the assignment marking through exercises with automated feedback (and a summative result for the instructor, see Figure 10: Quiz result, as viewed by instructor (small font = overview)) so that the students can benefit – apart from the automation giving them immediate corrective feedback – from the instructor working with them during class. Based on life experience, one would expect that she invests a portion of the saved assessment correction time into administrative tasks (I am not aware of her being research active), another portion into her leisure travels, and a final portion into he teaching, which could be during marking the remaining amount of (higher-level) assignments, occasionally glancing over the screenshot thumbnails in Sanako LAB 300 teacher interface and providing guidance to students having difficulty where such guidance cannot be automated by the exercise template).

The creation of thematic vocabulary exercises could be helped by the semantic NLP technology made available through Wordnet or Verbnet which both are included in the. E.g. "VerbNet is a hierarchical verb lexicon with syntactic and semantic information for English verbs, using Levin verb classes (Levin, 1993) to systematically construct lexical entries" (Kipper-Schuler:2005, p.28). Given the themes common in beginning to intermediate SLA introduction, it would be beneficial (ideally filtered by a frequency ranking for the subset appropriate for the current proficiency level) to replace all occurrences of verbs related to a theme, e.g. weather, by gaps.

```
>>> print verbnet.pprint('57')
weather-57
  Subclasses: (none)
  Members: blow clear drizzle fog freeze gust hail howl lightning mist
    mizzle pelt pour precipitate rain roar shower sleet snow spit spot
    sprinkle storm swelter teem thaw thunder
```

This requires

- extracting the set of thematically related words, as shown above (feature of the NLTK)
- restricting the set through frequency analysis:
  - NLTK
    - provides the function FreqDist(mylist) (NLTK-Book:2009, http://nltk.googlecode.com/svn/trunk/doc/book/ch01.html) which returns an ordered dictionary/hash containing the words and their number of occurrences, in frequency order.

        - NLTK also provides the datatype “ConditionalFreqDist, returning a “conditional frequency distribution [which] is a collection of frequency distributions, each one for a different ‘condition’. The condition will often be the category of the text.’(NLTK-Book:2009, http://nltk.googlecode.com/svn/trunk/doc/book/ch02.html)
        - Unfortunately, no stemming, let alone lemmatization of different word forms (inflections) out of the box. This functionality apparently would need to be added.
        - Also, insulated word frequency cannot identify the collocations (“bigrams that occur more often than we would expect based on the frequency of individual words”, NLTK-Book:2009: http://nltk.googlecode.com/svn/trunk/doc/book/ch01.html) most useful for language learners.
    - We can rely on the templates’ lookup features to automatically “gloss” rare and obscure words. Selecting movies based on the automatically difficulty (beyond frequency) of their subtitles would be yet another useful for SLA, but is beyond the scope of this paper.
- for the items passing the frequency requirement threshold, unfolding the inflections of each remaining member *counter-lemmatizing)

Since the exercise template is not very performant beyond 50 gaps, I divided the subtitles into chapters. This also made it easier to have more variation on the exercise topic: Each chapter could be devoted to a different (vocabulary or grammar) topic, by running regular expressions against each chapter text. In the future, it would be desirable to develop a feature which allows the combination (merging) or gap fill-creating (regular expressions or other) replacement operations, so as to be not too restrictive in testing.

3.2.2 PowerPoint Template for higher level (comprehension) language skills

In a second step, a PowerPoint Template to go through sequential question and answer as well as interactive exercises can be used whenever a segment of the movie has been viewed. Typically, both sequential and interactive templates are useful for focusing the student attention before viewing a segment and have them produce additional language during and after the viewing, versus adopting a passive “consumer” attitude which tends to happened when watching (especially not directly textbook accompanying) video in class. I have used the sequential templates to introduce vocabulary (and asked the students to jot down the context the new/difficult words were used in) or provide guiding comprehension questions (both with regards to “listening for the jist” or more specific information like dates and other numbers or crucial theses) which then gets revealed after the second column. In the *Miami Vice* example, one could e.g. ask for synonyms for “drugs” (see Figure 11: PPT view of teacher) and reveal these as they are provided by the class members (see Figure 12: PPT view of class).

3.2.3 Timeline Template for higher level (collaborative and comparative) plot-summary

I have also successfully used timeline-based exercises to have students interact with movie material. More powerful than a manual template which merely is meant to structure an individual students’ thought (see Figure 13: Timeline, fancy looking but non-automated) is a timeline Excel template which I have not created myself, but used in my teaching an timeline template as a basis for a structured collaborative and comparative plot summary and discussion (**Error! Reference source not found.**).

While naturally suited to the German Public TV series on cultural history of the 20th century called “100 deutsche Jahre”, which provided thematic cross-sections through the last century, revolving around topics like good manners, attitudes towards immigrants, emigration, wars, cars, vocational customs etc., it should also lend itself well to the both simple as slightly mystifying plot development of a televised *Whodunit?*

The basic premise is that a “shared”[16] MS-Excel file which contains on one sheet a timeline visualization of the structured data on a second sheet in is put on a student writable network share (as is part of the language lab environment[17]) for the students to edit after watching the sequel together in class,

In the worksheet 2 of the spreadsheet, there are rows, which will form the timeline x-axis, in one column of the workbook (for the “100 deutsche Jahre”, obviously 100 containing the years of the century; for the *Miami* Vice sequel possibly 45, one for each minute of the video which are also marked in the subtitle exercise file). Each class participant gets his own set of such rows with (years, minutes) in column A where she has to enter in column B notable turning points which then form a plot summary.

[16] See for this built-in and widely supported, but rarely used features REFERENCE: [Microsoft:] About shared workbooks; http://office.microsoft.com/en-us/excel/HP052622941033.aspx.

[17] Originally, the teacher collected – another language lab feature - regular workbooks from individual student computers and collated their text manually to feed the timeline visualization – the shared workbook approach is superior, especially if the student input is visualized in real time, as described below.

The task instruction provided to the students is to open the spreadsheet on worksheet 2 and enter around 10 notable events or other facts which can be dated (more apt for intermediate language learner classes than a more abstract discussion) from the '100 deutsche Jahre" and could be for the *Whodunit?* to enter the crucial turning points of the plot (from exposition with leading characters, "inciting moment/incident (Gustav Freytag: "erregendes Moment") to catharsis). The exercise design, focusing on short factual statements, focusing on basic and intermediate comprehension, reuse of new vocabulary as well as on form (for which the students would be well—advised to use the MS-Office proofing tools infrastructure), seems appropriate for intermediate proficiency work. At the same time, the exercise design can also be used to introduce the students to some narrative patterns and *topoi* (comic relief, romantic interest, subplots etc) which can commonly be found even in the lighter fare.

The Worksheet 1 timeline visualization will automatically collate the student contributions in an easily comparable format. The teacher on/from her computer can either monitor worksheet 2 and display the timeline sheet results to the class for discussion when everybody has finished the exercise, or already display the timeline visualization (which should refresh automatically) to the class during the execution of the exercise, which should make the exercise truly collaborative and (from my experience using text chat in beginner and intermediate classes) should entice more (non-spoken) communicative interaction than an intermediate class is normally capable of sustaining.

Since the workbook is shared, the teacher can intervene, including correcting students' language, either during exercise execution or summary discussion, for all to see and learn from in the timeline visualization, but without anybody being put "on the spot, since the contributions are anonymous.

This exercise type requires little extra work or preparation from the teacher, except familiarity with the plot (which probably can be sufficiently refreshed by quickly scanning the subtitles before class rather than having to watch again the sequel). However, if need be, the instructor can prepare the cells which the student fills out in worksheet2 with questions/fragments of statements for the students to answer/complete, calling their attention to turning points in the plot, segments which are difficult to understand or plot holes, like in this *Miami Vice* example (see Figure 15: Timeline Questions Miami Vice).

4 "technical limitations ", relieving the teachers from "the user-friendly (or unfriendly) interface between the material and the design" or helping them "get used to it"

4.1 Back office

The teachers would not deal with any of the back office operations - that is one of the value proposition of having a technical support person in the ELT department - let's hope each ESL department in the world will have at least one or sees the need for creating a position for one. The back office has to deal with sourcing subtitles, changing the time-in-/-out information to reformatting, programming the regular expressions and running them against the file to replace function words - and ideally in the future thematic vocabulary that falls within a certain frequency range. To be more precise:

What has not even been addressed is the need for sourcing DVDs with target language subtitles - few editions contain them -, and extracting the subtitles (which is a technical process as they are not included in a text format on the DVD). Alternatively, one can try and crowdsource this part of the job, i.e. hunt the subtitles down on the internet. The availability of subtitles on the internet seems to have much improved. Copyright also needs to be addressed, but one would assume the purchased of a DVD which includes the text in audio/spoken form, entitles one also to the use of the text version of this audio, no matter where sourced from.[18]

If one uses the shortcut of downloading subtitles rather than extracting the subtitles that came with the DVD, one has to make sure that the subtitles and (specific cut of the) movie editions/timeline actually match. DVD movies come in an amazing number of different (for different distribution channels, different audiences, including national market and other marketing reasons I do not understand) editions which are poorly documented (the field seems still in its infancy, online catalogues of DVD retailers are not very useful in this respect, and library science does not go there, at least not very far). While a cut or shortened scene does not have as devastating an effect with our way to distribute subtitle exercises on the side compared with displaying them as part of the movie window, differences between subtitles and actual movie can confuste the language learner. To massage the subtitle time-in/out information is the purpose of the (admittedly rather convoluted) MS-Excel database of subtitles (Figure 3: Subtitles converted to XLS).

Only after the back office has converted the subtitles into a more legible format, programmed the regular expression and run it against the subtitles text to mark it up for the MS-Wod template macro to generate a gap-fill quiz, the teachers would come into play only (apart from providing ideas which functionwords and thematic vocabulary to replace). The teacher would receive completed MS-Word based exercise files (one or multiple per movie, depending on text size, maybe by "chapter", as commonly contained in DVD movie editions) for distribution, and again when the studends submit the MS-Word exercises

[18] I would expect the entertainment industry to worry much more about the cavalier treatment of film rights primarily in the wider public, secondarily in many SLA departments I know.

they completed. The teacher only needs top open up the submissions (ideally from Windows explorer, take in the summary information from the opening dialogue, click “OK” on each dialogue, then close all files/exit word).

Only if the teacher does not want to limit the exercises to what can be automated, she can also receive an execise template with plain subtitle (chapter) text and mark the portions of the text to be omitted herself (training for that is in the screencast quiz_create_test_or_take).

4.2 Teacher training

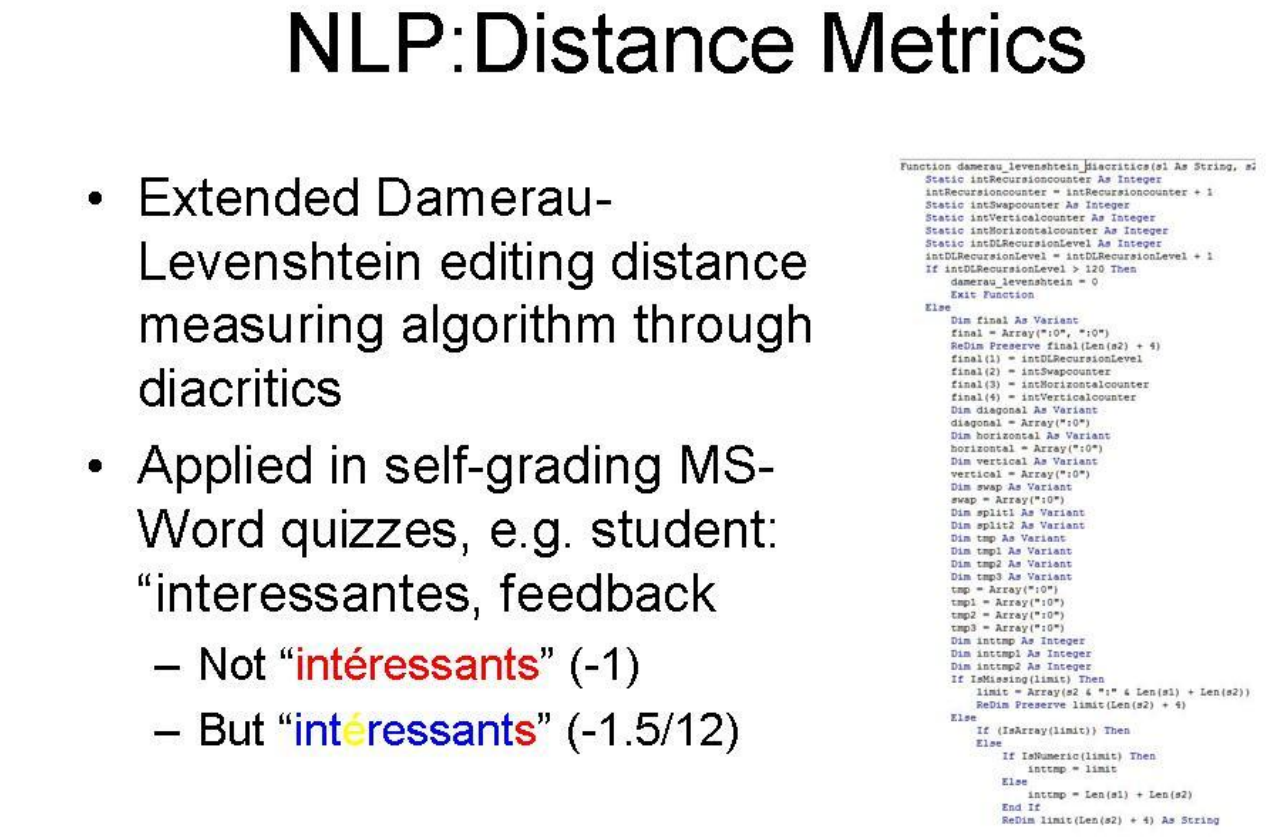
I also made a cheatsheet – for the teachers to take home from a workshop and put next to their desktop screen - to explain on a single page all that is needed to know when using the face-to-face teaching templates (see Figure 16: Teacher Training Cheatsheet).

However, I have had a lot of issues with teachers not learning computer essentials for knowledge workers since 1991, including what is a cheatsheet and why do I need to attend a training in the first place. These are issues beyond the scope of this paper (as well as my roles in e-learning beyond this paper). After all, I can not give up the hope to expect a rational metric being implemented in SLA and consequently, that ESL teachers will be willing to put up with the current limitations of computer technology I order to reap its benefits (although I acknowledge that SLA teachers’ apparent reluctance may be based on superior insight by, in the race between their own evolutionary development and CS approaching “the Singularity”, betting on the latter (thus transcending the former?) to technically realize the “artificial secretary/clerk/servant/exploited adjunct or grad student” U – as opposed to Tainter:1988, p 123ff., seeing technological innovation only as a temporarily delaying the decline of marginal return on investment).

5 Outlook

- SLA to-do’s
  - I am finished with a Damerau-Levenshtein distance metrics[19] implementation in VBA, but have not hooked it up to the quiz.dot. While distance metrics is mechanic and nearly language agnostic, it could relieve (by intelligence amplification) even more an instructor in a face-to-face teaching situation where the instructor

[19] See http://en.wikipedia.org/wiki/Damerau%E2%80%93Levenshtein_distance, example here:



monitors a multitude of student quizzes in thumbnail view, but also for the student doing these exercises, if slight differences like spelling errors or even only erroneous diacritics, could be indicated in a different color than wrong words chosen.

- Other possible choices
  - Numbers: “Much/many” dichotomy
  - “Which/who” relative pronoun dichotomy: Difficult for German students of English which has no such dichotomy for innate beings/things, but whose (antiquated) relative pronoun “Welch” as a false friend of which tends to lead to a wrong preference of “which” to “who” by German speakers.[20]
  - Sub clauses/tenses: if clauses up to period/comma, giving the number of words as hints. Would require a delegate.
- More advanced replacements which, beyond regular expressions, rely on POS tagging, and integation of POS tagging tools form the field of corpus linguistics

- “plumbing” To-do’s
  - Script should allow for passing multiple (longer movies split into installments or sequels of TV series) files from the command line (like “Miami Vice*.srt” processing all subtitle files for so many seasons of the once popular TV series).

[20] I know at least one such speaker personally.

- Generally speaking, these applications of “artificial intelligence” to language teaching tasks remains very low-level, mechanic and almost natural language agnostic. However, it seems the incredibly low cost of these operations could result in a quantity that has a qualitative impact on SLA. Especially if theses tasks the computer can do are mundane, the human instructor should be relieved of them to be able to focus on higher intelligence tasks which the computer can not do. We may use as a model the first forms of non-artificial intelligence (move to close to a volcanic spring = burn, move too far away = freeze and starve) , while primitive, managed set of a massive reaction to feedback data cycle which can grow into something more impressive over time.

6 Appendix

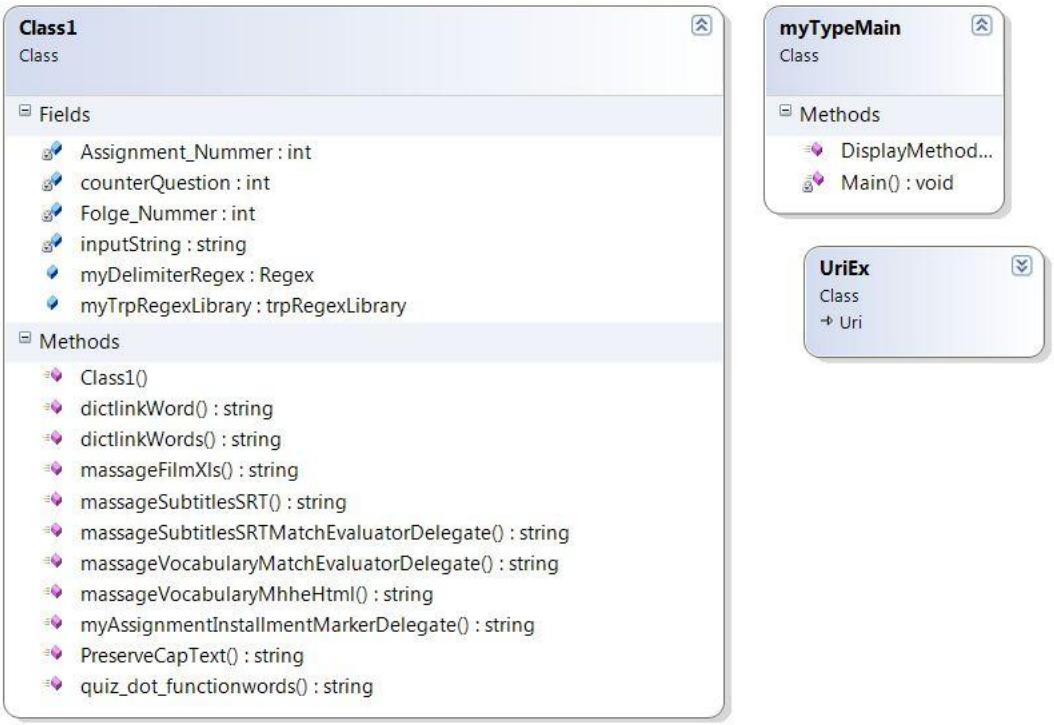


Figure 1: Vs.net project class diagram

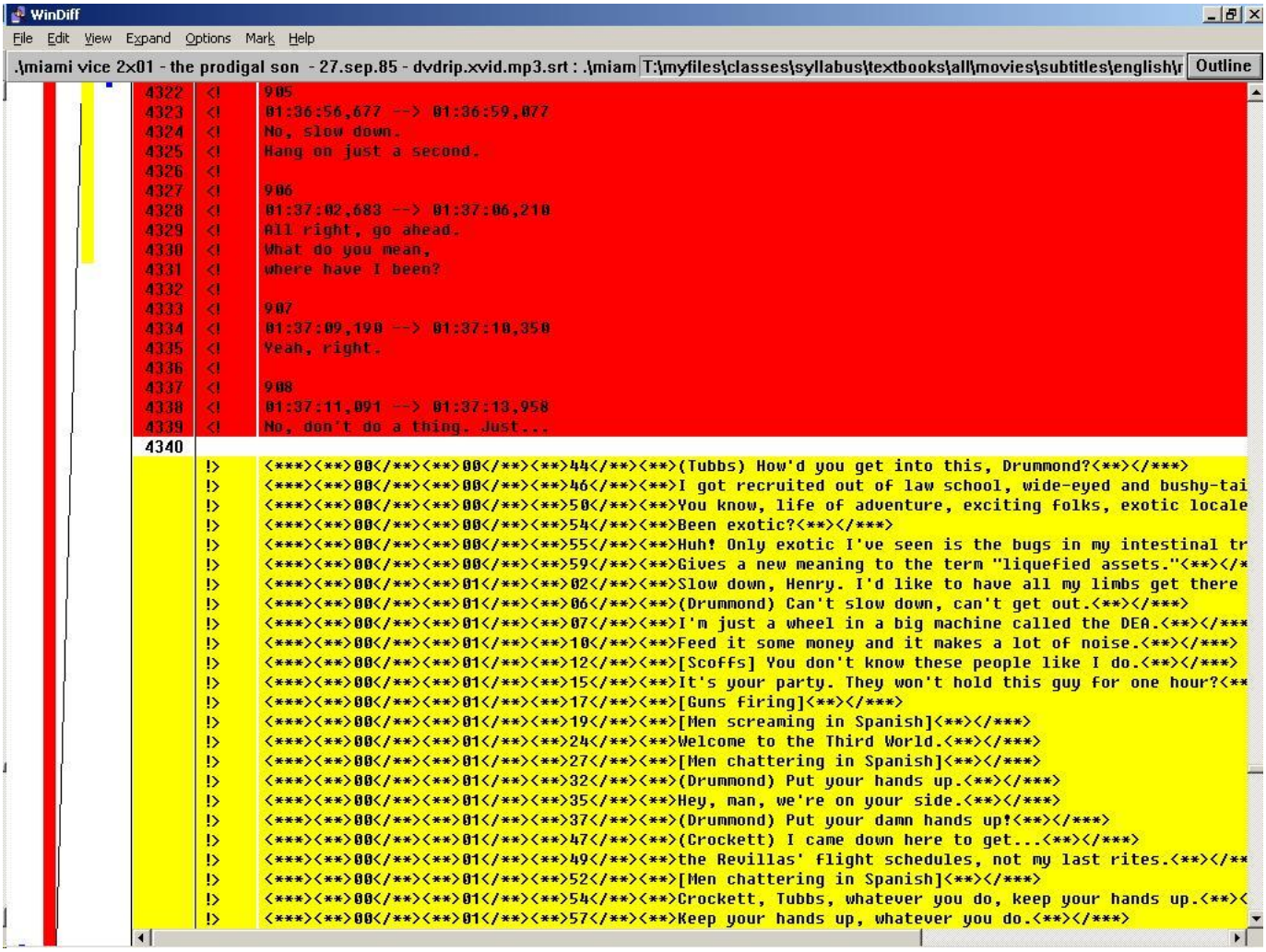


Figure 2: Windiff of MassageSubtitlesSRT

5	general			ancillary		dvd (target)					subtitles (source)						
6	lang	title	apter, custom subdiv.	earity	curr delta, empirical & terpolated	sec	hh	mr	as minutes	ss	se	hh	mr	ss	ce	subtitle
	rt Ascending	Vita è bella		0	0	35	0	0	0	35	35	0	0	35		*Questa é una storia semplice,*
	rt Descending	Vita è bella		0	0	38	0	0	0	38	38	0	0	38		*non é facile raccontala.*
	l)	Vita è bella		0	0	40	0	0	0	40	40	0	0	40		*Come in una favola, c'é dolore*
	op 10...)	Vita è bella		0	0	43	0	0	0	43	43	0	0	43		*ma é anche piena di meraviglie e di felicità.*
	ustom...)	Vita è bella		0	0	50	0	0	0	50	50	0	0	50		Canto quello che vedo su niente soprassiedo.
	rman	Vita è bella		0	0	53	0	0	0	53	53	0	0	53		Eccomi, dissi al caos. "Son tuo schiavo".
	lian	Vita è bella		0	0	56	0	0	0	56	56	0	0	56		Egli rispose:"Bravo"! "Di che", dissi io "Chi fu?"
	anish	Vita è bella		0	0	60	0	1	1	0	60	0	1	0		Libero alfine sono. Che serve una carezza quando
	anks)	La Vita è bella		0	0	64	0	1	1	4	64	0	1	4		di me l'ebrezza s'impossessó?
	onBlanks)	La Vita è bella		0	0	65	0	1	1	5	65	0	1	5		Eccomi sono pronto,
2779	italian	La Vita è bella		0	0	67	0	1	1	7	67	0	1	7		Son già partiti i treni, si sono rotti i freni,
2780	italian	La Vita è bella		0	0	70	0	1	1	10	70	0	1	10		e non resisto più. Vai, dolce Bacco, portami...
2781	italian	La Vita è bella		0	0	73	0	1	1	13	73	0	1	13		Si sono rotti i freni! Si sono rotti i freni!
2782	italian	La Vita è bella		0	0	76	0	1	1	16	76	0	1	16		Ho capito!
2783	italian	La Vita è bella		0	0	78	0	1	1	18	78	0	1	18		- No, si sono rotti per davvero! - Frena!
2784	italian	La Vita è bella		0	0	81	0	1	1	21	81	00	01	21		Ma non stavi recitando la poesia? Stai attento ci sono i lup
2785	italian	La Vita è bella		0	0	82	0	1	1	22	82	00	01	22		Non funziona!
2786	italian	La Vita è bella		0	0	90	0	1	1	30	90	00	01	30		Frena!
2787	italian	La Vita è bella		0	0	92	0	1	1	32	92	00	01	32		Ci si ammazza! C'é il basco!
2788	italian	La Vita è bella		0	0	100	0	1	1	40	100	00	01	40		Frena!
2789	italian	La Vita è bella		0	0	107	0	1	1	47	107	00	01	47		Sta arrivando il re!
2790	italian	La Vita è bella		0	0	109	0	1	1	49	109	00	01	49		Eccolo!
2791	italian	La Vita è bella		0	0	112	0	1	1	52	112	00	01	52		Gira, c'é un monte di gente laggiù! Diritto!
2792	italian	La Vita è bella		0	0	115	0	1	1	55	115	00	01	55		Non ti muovere non ci sono i freni!
2793	italian	La Vita è bella		0	0	117	0	1	1	57	117	00	01	57		Non ci sono i freni, via, viaaa!
2794	italian	La Vita è bella		0	0	121	0	2	2	1	121	00	02	01		Via! Via!
2795	italian	La Vita è bella		0	0	130	0	2	2	10	130	00	02	10		Viva il re! Viva il re!
2796	italian	La Vita è bella		0	0	146	0	2	2	26	146	00	02	26		Fatti una passeggiata, dieci minuti
2797	italian	La Vita è bella		0	0	148	0	2	2	28	148	00	02	28		Sennò arriviamo domattina.
2798	italian	La Vita è bella		0	0	150	0	2	2	30	150	00	02	30		Ho trovato il chiodo di prima, lo vuoi?
2799	italian	La Vita è bella		0	0	152	0	2	2	32	152	00	02	32		Grazie, dammelo, vai...
2800	italian	La Vita è bella		0	0	154	0	2	2	34	154	00	02	34		Ora che ti do?

Figure 3: Subtitles converted to XLS

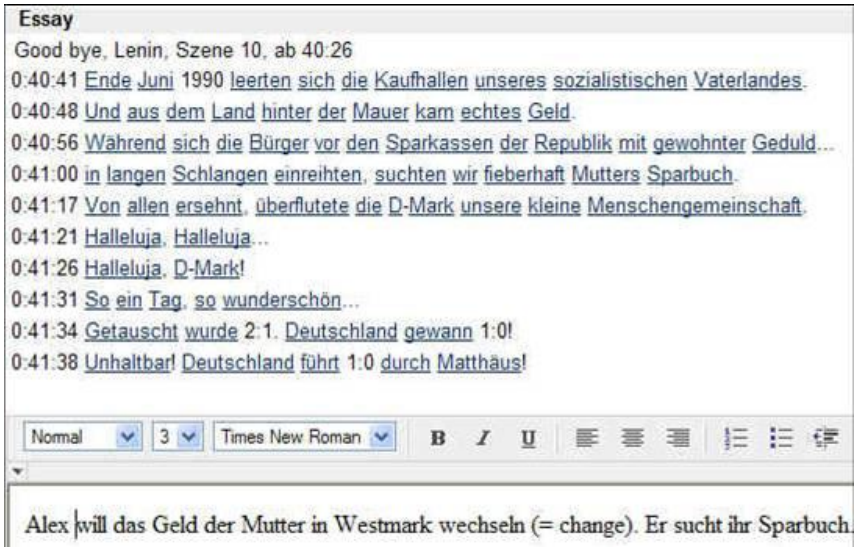


Figure 4: Glossed Subtitle VLE Assessment, Student answering

Essay **100 points**

Question Good bye, Lenin, Szene 10, ab 40:26

0:40:41 Ende Juni 1990 leerten sich die Kaufhallen unseres sozialistischen Vaterlandes.

0:40:48 Und aus dem Land hinter der Mauer kam echtes Geld.

0:40:56 Während sich die Bürger vor den Sparkassen der Republik mit gewohnter Geduld...

0:41:00 in langen Schlangen einreihten, suchten wir fieberhaft Mutters Sparbuch.

0:41:17 Von allen ersehnt, überflutete die D-Mark unsere kleine Menschengemeinschaft.

0:41:21 Halleluja, Halleluja...

0:41:26 Halleluja, D-Mark!

0:41:31 So ein Tag, so wunderschön...

0:41:34 Getauscht wurde 2:1. Deutschland gewann 1:0!

0:41:38 Unhaltbar! Deutschland führt 1:0 durch Matthäus!

Answer Alex im Supermarkt vor leeren Regalen. Suche nach Mutters Sparbuch. Geldtransporter bringen DM, das neue Geld wird gefeiert

Figure 5: Glossed Subtitle VLE Assessment, Instructor model answer

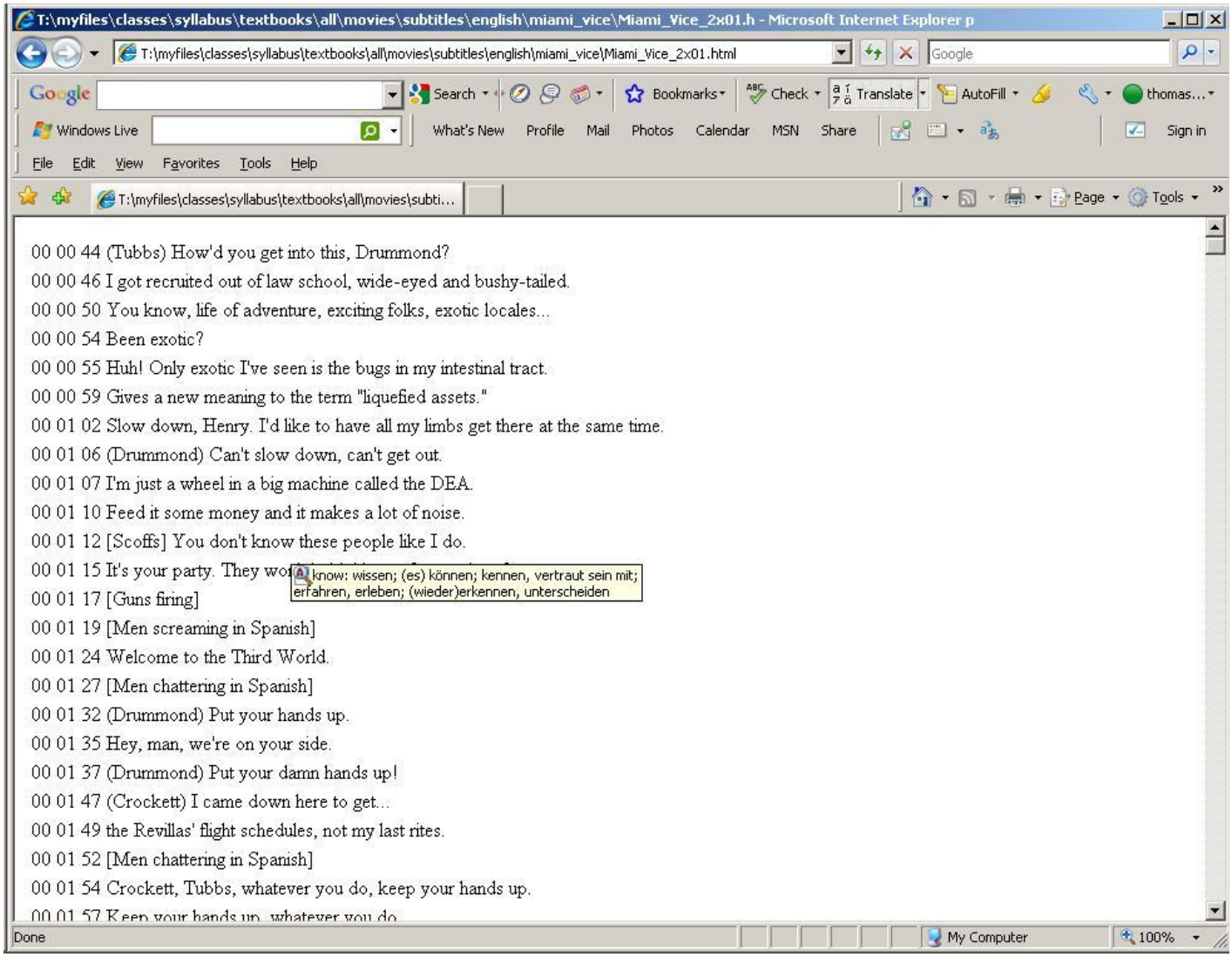


Figure 6: Subtitles Reformatted as HTML Table

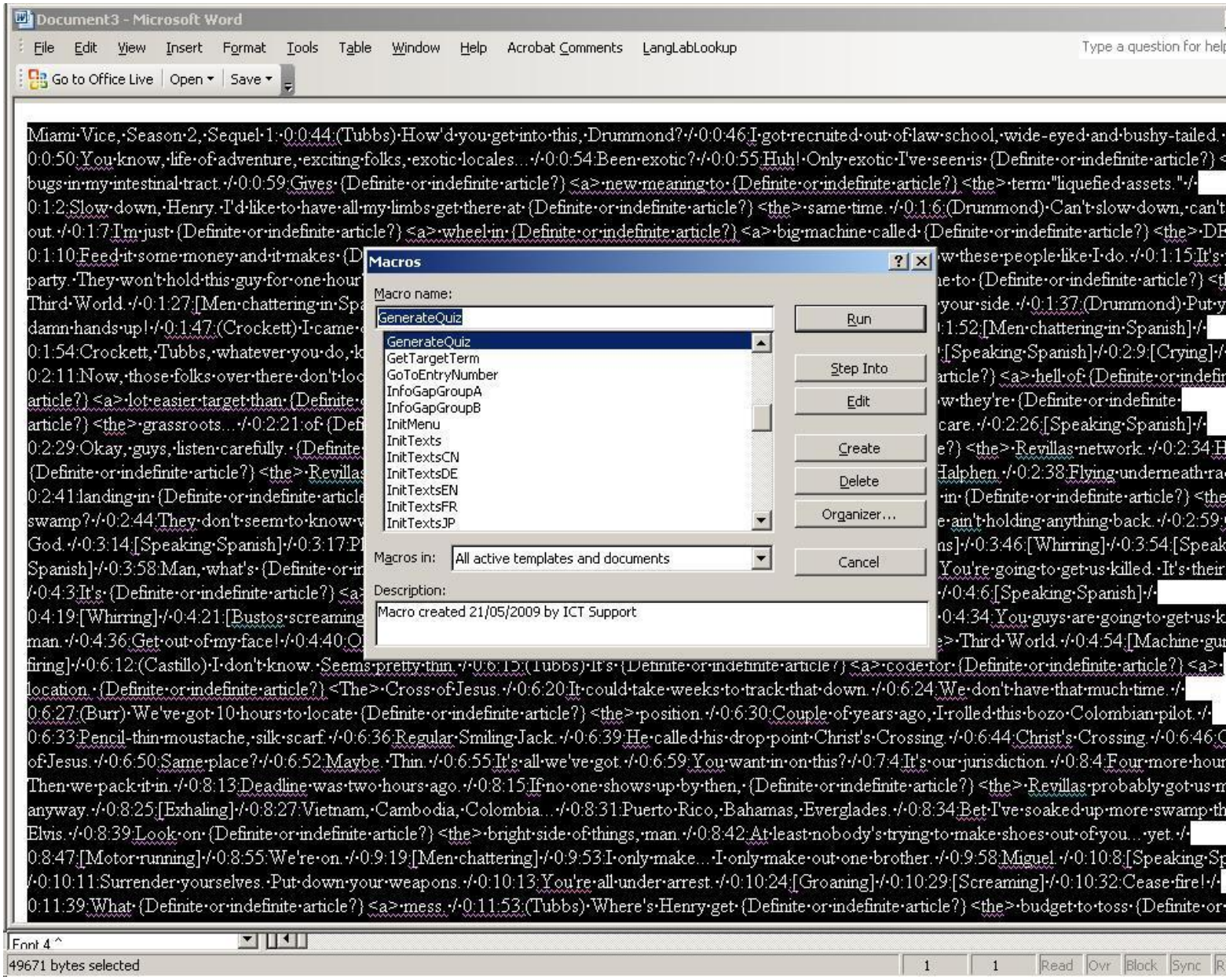


Figure 7: subtitles ready for quiz generation in quiz.dot

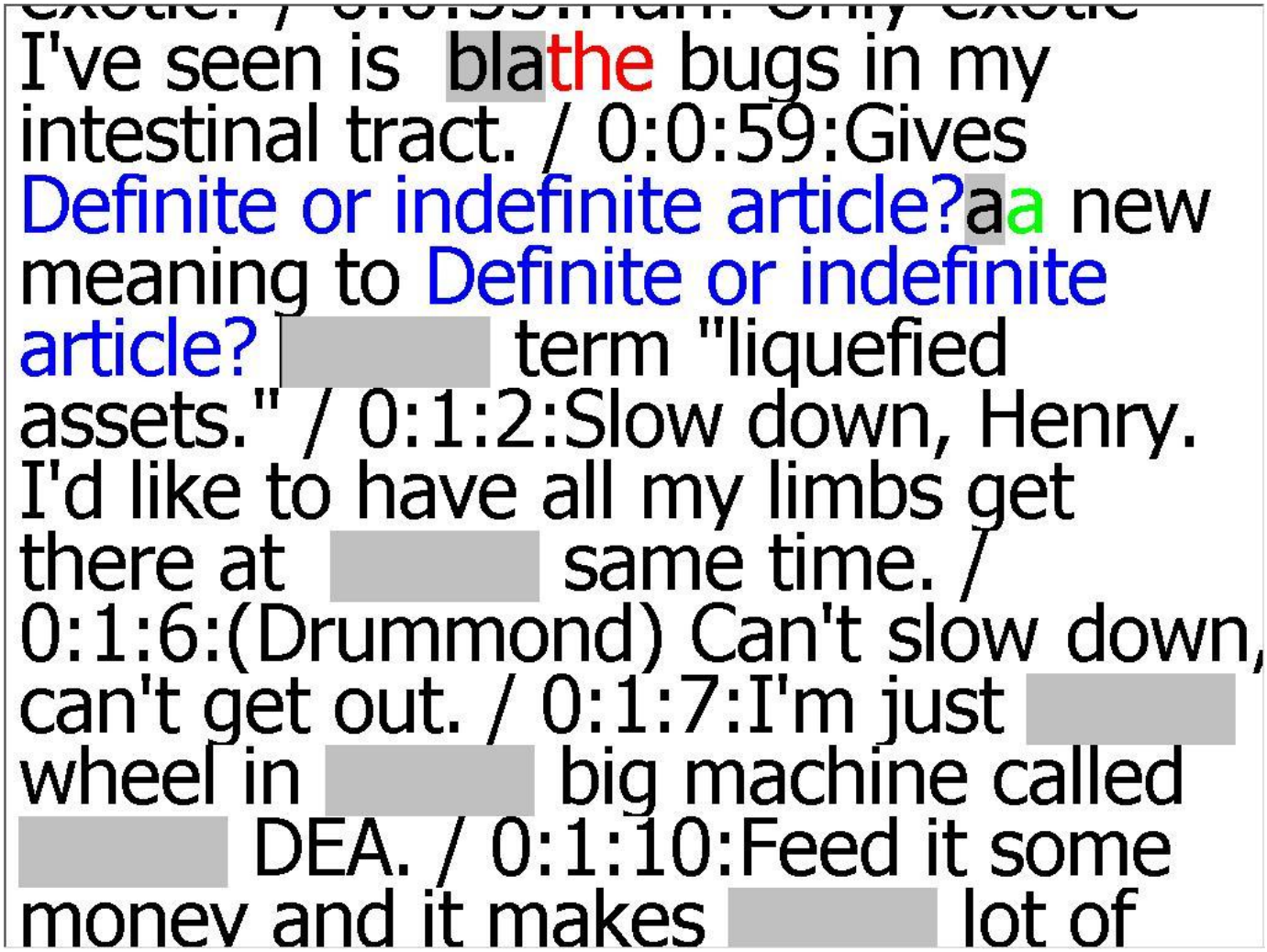


Figure 8: Quiz as viewed by Student

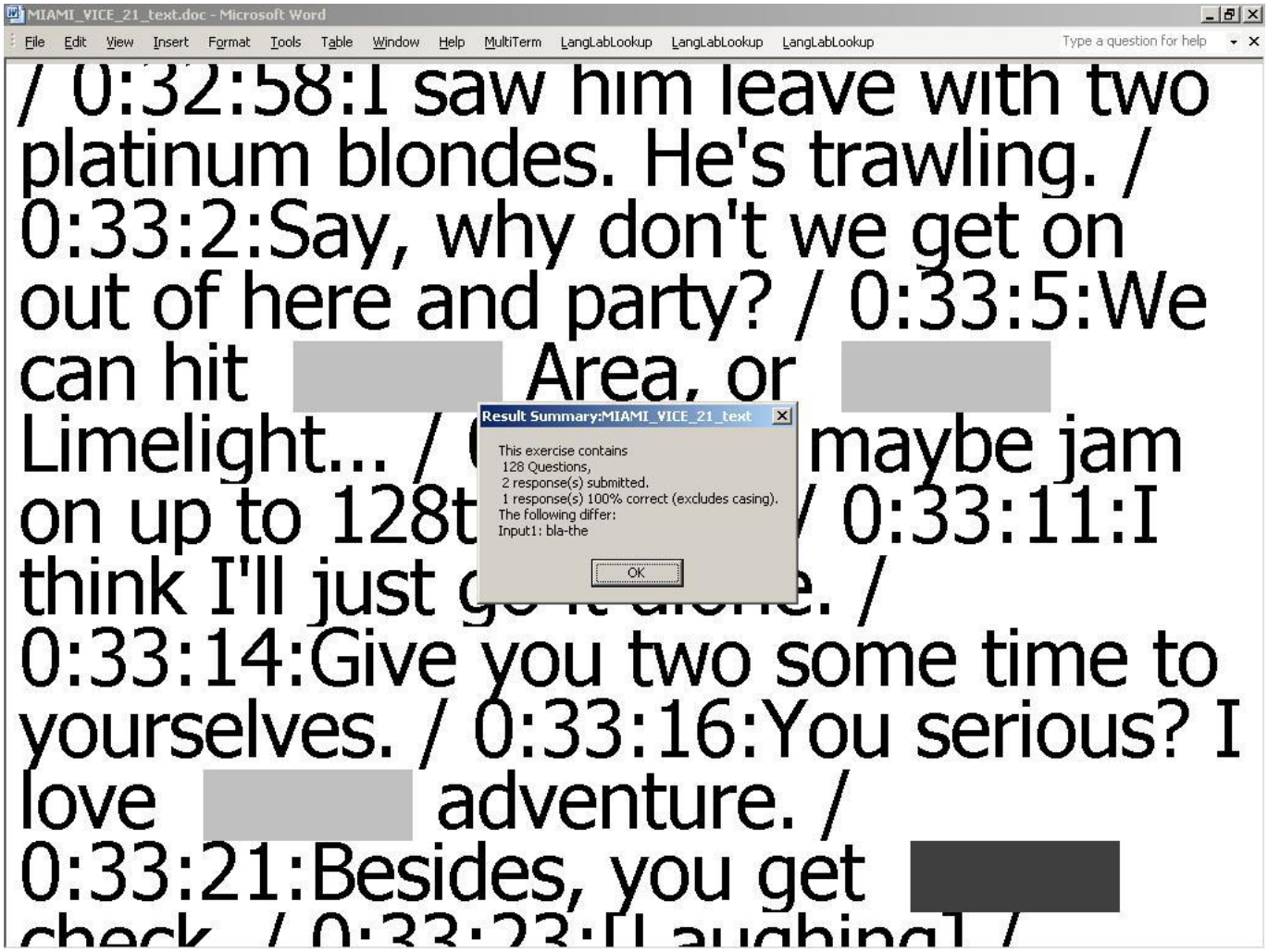


Figure 9: Quiz result summary, as viewed by student

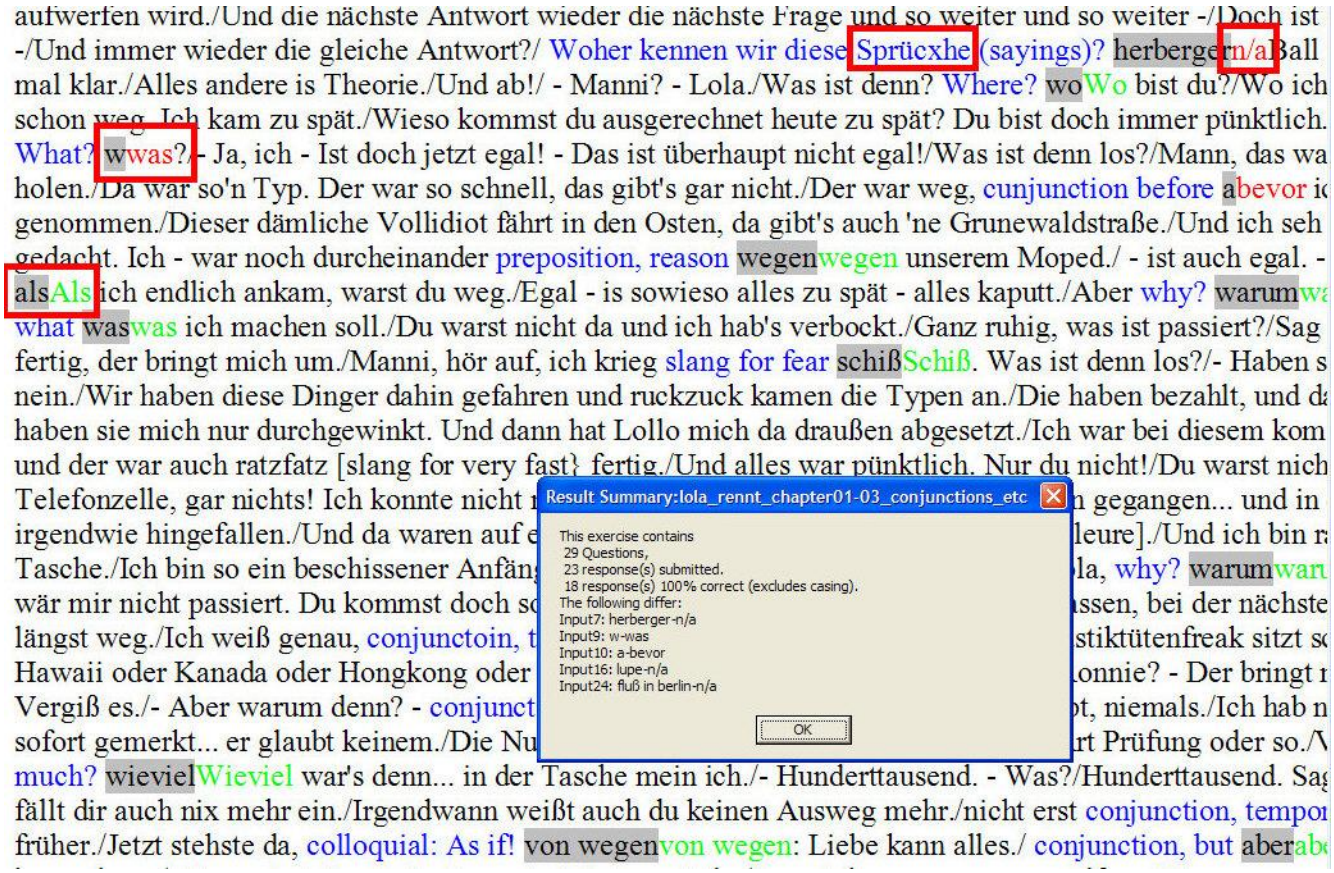


Figure 10: Quiz result, as viewed by instructor (small font = overview)

Figure 11: PPT view of teacher

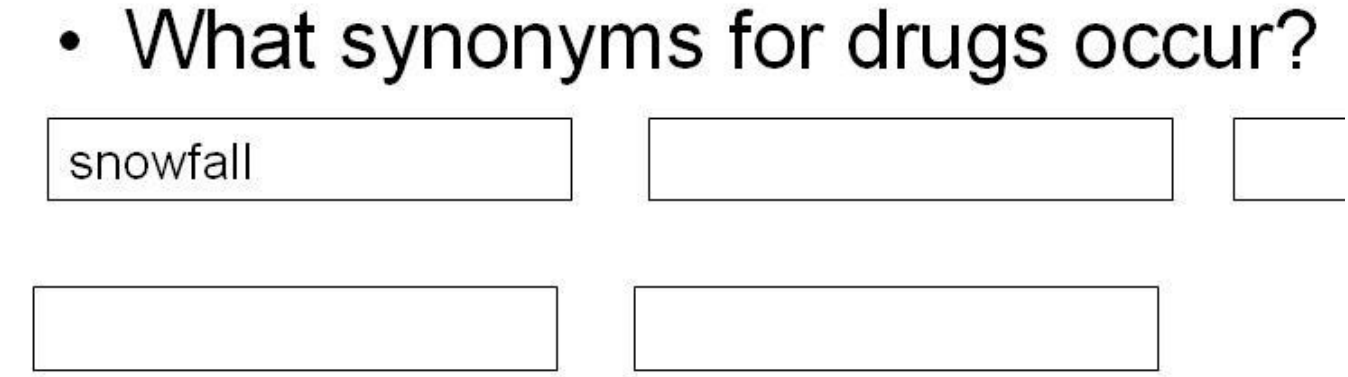


Figure 12: PPT view of class

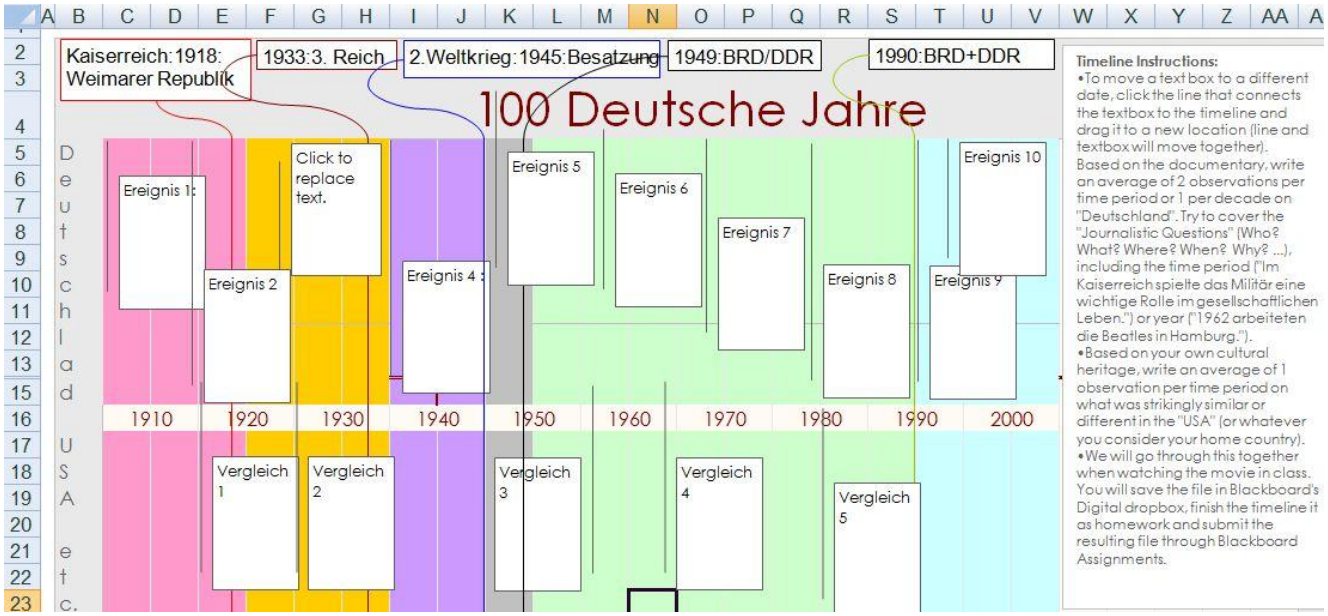


Figure 13: Timeline, fancy looking but non-automated

Title: 100 Deutsche Jahre

Eventvalue	Authorvalue	Author	Year	Event	Height	Axis
87	5	larry	1900	Viele Deutsche emigrierten in die USA, weil sie dort ein	-30	1900
72	5	larry	1900	Die Polen, die keine Arbeit haben, arbeiten in den deuts	0	1900
67	6	thomas	1900	Völkerschauen unterhalten die Deutschen mit exotischer	20	1900
63	6	thomas	1910	Eine halbe Million Ruhrpolen arbeiten in Deutschland unt	-10	1910
23	3	trp	1918	Polen wird neugegründet	10	1918
55	3	trp	1920	Die Polen in Deutschland assimilieren sich (Fußball...).	-30	1920
46	6	thomas	1935	Italiener bauen die Volkswagen-Stadt Wolfsburg	-10	1935
45	6	thomas	1945	8 Millionen Fremdarbeiter sind in Deutschland	10	1945
64	3	trp	1955	Das erste Gastarbeiterabkommen (mit Italien) wird unte	20	1955
98	5	larry	1961	Viele Ostdeutsche ziehen nach Westdeutschland um, de	30	1961
101	5	larry	1961	die Sowjetunion baut die Berliner Mauer, denn (for) viel	-30	1961
102	5	larry	1961	die Sowjetunion baut die Berliner Mauer, weil (since) vi	-20	1961
93	5	larry	1961	weil viele Ostdeutsche nach Westdeutschland umziehen,	-10	1961
44	5	larry	1961	1961 baut die sowjetunion die Berliner Mauer	0	1961
73	5	larry	1961	Die Sowjetunion, die viele Arbeiter verloren hat, baut di	10	1961
70	6	thomas	1961	Die Berliner Mauer wird errichtet. Keine Flüchtlinge aus	20	1961
71	3	trp	1965	DDR bildet Vietnamesen aus, zur Hilfe im Kampf gegen d	30	1965
58	6	thomas	1973	Anwerbestopp für Gastarbeiter (Ölkrise, Wirtschadftskris	-25	1973
31	3	trp	1980	Vietnamesen arbeiten in der DDR	-15	1980
25	3	trp	1991	Rostocker Ausschreitungen	-5	1991
31	3	trp	1992	Ausschreitungen in Mölln und ?	5	1992
43	3	trp	1996	BRD reformiert und beschränkt das Asylrecht	10	1996

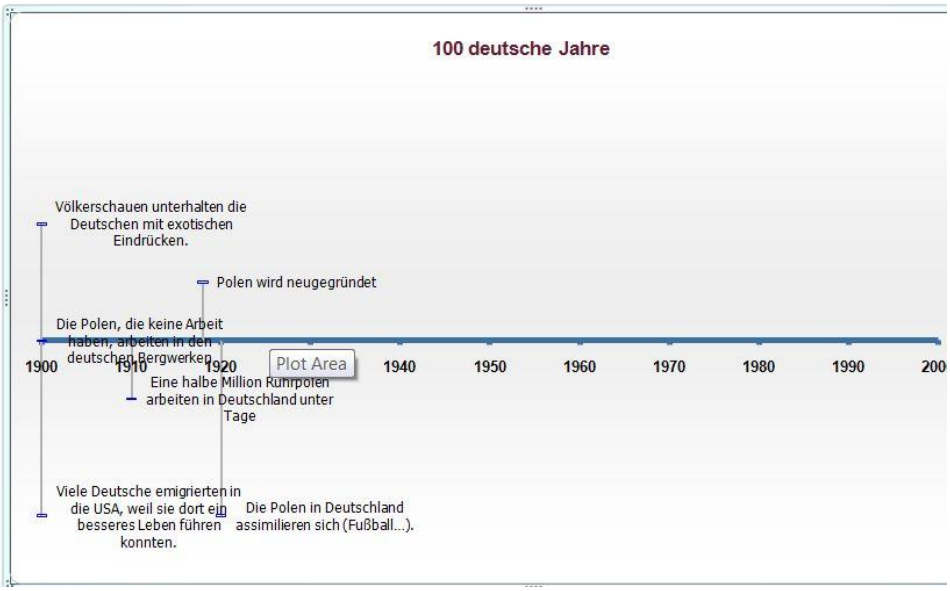


Figure 14: Timeline Template for Comparative/Collaborative Plot Summary Exercise

00:44	Where are our agents?
01:49	Who are the revillas?
02:34	How do the Revillas smuggle drugs?
13:06	who is gina
14:22	What happens after the success in the swamp?
17:42	Who is Newton Windsor Blade?
19:33	Who is Jimmy Borges?
21:27	Who is Valerie?
28:27:00	Who is Sacco?
38:58:00	Who is Margaret?
82:13:00	What role does Wall street play in the drug smuggling?

Figure 15: Timeline Questions Miami Vice

authoring a word quiz	doubleclick K:\centre\configuration\microsoft\templates\student_writing_quiz.dot
	enable macros
	menu:tools/unprotect
	type or copy/paste text
	enclose answers in {}
	type hints and enclose <>
	set tools/ language
	menu:tools/macros/macros/select generatequiz, run
	review summary
	not happy? undoQuiz
	repeat until done
	menu:file/close
	save on k:\teachingmaterials (students can read)
taking a word quiz	enable macros in doc launch to you by teacher
	tab into formfield
	study hint, if any
	type response into formfield
	tab out to reveal feedback
	if scrolled, scroll back
	study feedback
	repeat until done
	use lookup as needed
	menu:file /close
	save / collect [in study1200 ver2 still manually]
	study summary feedback
grading a word quiz	open student saved file
	study summary feedback
taking word notes	enable macros in doc launched to you by teacher
	write
	use lookup as needed
	menu:file /close
	save
creating a ppt quiz	doubleclick template M:\centre\configuration\microsoft\templates\teacher.pot
	in slides overview, context-menu:copy/paste
	replace placeholder text
	delete unneeded placeholders
	test with shift-f5,right-arrow through sequence/mouse-click on interactive parts
using lookup	if quiz lookup menu / pause quiz
	if quiz, doubleclick word
	select word/phrase, choose from lookup menu
	if writing, copy/paste info as notes

Figure 16: Teacher Training Cheatsheet

7 References

Kipper-Schuler:2005 : Karin Kipper-Schuler VERBNET: A BROAD-COVERAGE, COMPREHENSIVE VERB LEXICON. Diss. University of Pennsylvania 2005.

Kipper-Schuler:2009: Karin Kipper-Schuler*: VerbNet Lexicon, Version 2.1* [download | source] .

NLTK-Book:2009: Steven Bird, Ewan Klein, and Edward Loper: Natural Language Processing with Python --- Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009 (also at http://nltk.googlecode.com/svn/trunk/doc/book/)

Tainter:1988:Joseph A. Tainter:The Collapse of Complex Societies. New York 1988