Towards Multimodal Spoken Language Corpora: TransTool and SyncTool

Joakim Nivre Jens Allwood Jenny Holm Dario Lopez-Kästen
Kristina Tullgren Elisabeth Ahlsén Leif Grönqvist Sylvana Sofkova

Göteborg University
Department of Linguistics
E-mail: `firstname.lastname@ling.gu.se`

Abstract

This paper argues for the usefulness of multimodal spoken language corpora and specifies components of a platform for the creation, maintenance and exploitation of such corpora. Two of the components, which have already been implemented as prototypes, are described in more detail: TransTool and SyncTool. TransTool is a transcription editor meant to facilitate and partially automate the task of a human transcriber, while SyncTool is a tool for aligning the resulting transcriptions with a digitized audio and video recording in order to allow synchronized presentation of different representations (e.g., text, audio, video, acoustic analysis). Finally, a brief comparison is made between these tools and other programs developed for similar purposes.

1. Introduction

The availability of adequate tools for the creation, maintenance and use of multimodal spoken language corpora is an important instrumental goal for spoken language research, whether this research is motivated primarily by the desire to gain a better understanding of the mechanisms of spoken communication or by the wish to develop practical applications such as multimodal interfaces for human-machine interaction.

Multimodal dialog systems will be a feature of many future applications, e.g., information systems. They will also be a feature of many VR systems and tutoring systems. The basic source of inspiration for dialog systems is ordinary human face-to-face communication involving both speech and gestures. However, our understanding of human communication as a multimodal phenomenon is still very insufficient. Thus, there is a need for tools which will enable us to gain a better understanding of the relations between properties of human face-to-face communication, such as gestures, intonation, words and grammar, and of how the utterances and gestures of different speakers are coordinated with each other.

In this paper, we report on a long-term project to develop a platform for multimodal spoken language corpora. More specifically, we describe two modules of such a platform, which both exist in prototype implementations. The first of these modules, which is called TransTool, is a transcription editor which assists a human transcriber in producing transcriptions in accordance with a given standard and partially automates some of the tasks involved, e.g., in the marking of overlapping speech.

The second one, SyncTool, is a tool for aligning transcriptions with the corresponding digitized audio and video recordings in order to allow synchronized display of different representations. Again, this is meant to provide support for a human analyst rather than to provide a completely automated process, although the latter would of course be preferable in the long run.

Before we turn to a detailed description of TransTool and SyncTool, however, we will try to set the stage by presenting the platform for multimodal spoken language corpora of which these tools are meant to be part.

2. Background

Even though face-to-face spoken interaction is one of the most basic forms of human communication, many aspects of it are still not well understood. To some extent, this lack of knowledge is due to a lack of good data as well as adequate tools for presenting and analysing the data. In order to study spoken communication efficiently, we need not only recordings of naturally occurring spoken communication and transcriptions of these recordings, but also tools for presenting and analysing these transcriptions and recordings in a flexible and efficient manner.

The picture is further complicated by the fact that face-to-face spoken communication is multimodal, involving gestures as well as speech, which means that video recordings are usually required. But this also means that transcriptions must be synchronized and displayed together not only with an acoustic signal but also in conjunction with visual data on gestures, etc., which tends to magnify the technical difficulties involved.

Putting together a multimodal spoken language corpus is a very labor intensive task. First of all, manual transcription is laborious and time-consuming in itself, and even more so if the multimodal aspects of spoken communication are taken into consideration. To this we have to add the work needed to insure that transcriptions and recordings are properly aligned, so that they can be displayed and analysed in a synchronized fashion.

In order to improve the situation, we need to develop adequate tools for the creation, maintenance and exploitation of multimodal spoken language corpora. Wherever possible, the aim should be to automate the processing, but for many of the tasks involved we will probably have to be content with providing computer support for manual work, support which either speeds up the process, or reduces the error rate, or indeed both.

The Department of Linguistics at Göteborg University has been involved in the empirical study of face-to-face spoken communication since the late 1970s. This has resulted in a corpus of transcribed spoken Swedish, which contains a wide variety of different activity types and which currently contains about one million word tokens (cf. Allwood 1998).

Over the years, a fairly detailed transcription standard has been developed, the most important features of which are the following (cf. Nivre 1998).

- A transcription consists of a *header*, containing background information such as type of activity, participants, date of recording, duration, transcriber, etc., and a *body*, containing the transcription proper.
- The transcription body consists of *speech lines*, containing the transcribed speech of dialog participants (each line introduced by a speaker initial); *comment lines*, containing comments pertaining to phenomena in speech lines (see below); *section lines*, indicating boundaries between subactivities or topics; and *time lines*, containing information about the amount of time elapsed from the start of the activity.
- Words are transcribed using standard orthography modified to capture spoken language variants and reductions (e.g., the Swedish first person pronoun 'jag' is transcribed 'ja' or 'jag' according to the actual pronunciation).
- Spoken language variants are indexed for disambiguation to the level of standard orthography (e.g., the Swedish first person pronoun is transcribed 'ja1' to distinguish it from the affirmative particle 'ja0' [yes]).
- Overlapping speech is marked by means of indexed square brackets (where the same index on different pairs of brackets indicate simultaneity).
- Comments are made by enclosing the commented part of an utterance in (possibly indexed) angle brackets and putting the actual comment in matching brackets on a separate comment line. An elaborate system of standardized comments, including comments for the coding of gestures, allows automatic parsing of comment information.

A short extract from a transcription, exemplifying most of the features discussed above, can be found in Figure 1.

Producing transcriptions in accordance with this standard is a very time-consuming task without adequate tools. It can also be an error-prone task, mostly because it involves a lot of numerical indexing (of words, comments, overlaps, etc.). This is the reason that we have thought it necessary to develop a computer tool to assist the manual transcriber in this process and partially automate some of the tasks involved (cf. section 3).

However, although transcriptions of this kind constitute a useful form of data for the study of spoken language, they are nevertheless insufficient in themselves and need to be supplemented with the actual sound and video recordings. Moreover, transcriptions and recordings need to be aligned so that analysts can view them together and do various types of coding and analysis based both on the recording and the transcription.

In order to satisfy these needs, we believe that several tools are needed. Hence, we have embarked on the project of building a platform for multimodal spoken language corpora, consisting of a set of integrated tools for the creation, maintenance and use of such corpora. The planned tools of the platform are the following:

- A tool for digitizing audiovisual corpus data (recordings).
- A tool for producing and checking standardized transcriptions (TransTool).
- A tool for semi-automatic alignment of audio/video and transcribed text (SyncTool).
- A multimodal corpus presentation tool, allowing simultaneous and synchronized display of transcriptions and audio/video recordings.
- A transcription coding tool, including display of transcriptions in different formats, with optional use of synchronized audio/video display.
- An analysis tool for processing information from the coding tool (and from the corpus itself).

If possible, all tools should be implemented in a platform-independent way and preferably allow access via the Internet.

Before we go on to describe the two tools relating directly to transcription — TransTool and SyncTool — it might be worth addressing the question of why we have chosen to develop our own tools instead of using existing ones. The simple answer is that we have not so far been able to find any tools that provide the right kind of functionality in the right kind of environment. First of all, there is no abundance of software in this domain. Secondly, many of the programs that do exist are developed for a specific purpose or a specific standard, which makes them hard to use in other contexts. Finally, most of the programs are available only on one or two software platforms, which may or may not be a problem depending on whether this happens to coincide with the platforms that you are working with yourself.

However, although we have not so far been able to reuse existing tools, it is clearly important to be open to developments within the area. In section 5, we will therefore make some brief comparisons between our tools and similar programs developed by others. Hopefully, this can contribute to a better understanding of the similarities and differences between different approaches and pave the way for cooperation in the future.

```
$A: ja0 de0 e0 <14 [4 havsströmmarna ]4 som gör >14 att de0 e0
    andra förhållanden där borta
@ <14 gesture: B nods >
$B: [4 m0 ]4
```

Figure 1. Transcription extract
[Translation of A's utterance: 'yes it is because of the sea currents that there are other conditions over there'.]

3. TransTool

TransTool is a computer tool for transcribing spoken language in accordance with the standard developed within the research program Semantics and Spoken Language at Göteborg University, Department of Linguistics, and described in Nivre (1998) (cf. section 2).

The current implementation of TransTool is done in Tcl/Tk (Tool Command Language/ Toolkit) and runs (at least) in Unix, Macintosh and Windows environments. The latest version of the program can be downloaded from http://www.ling.gu.se/gmslc/.

TransTool is equipped with File-, Edit- and Format menus which operate in much the same way as in word processing programs (Figure 2).

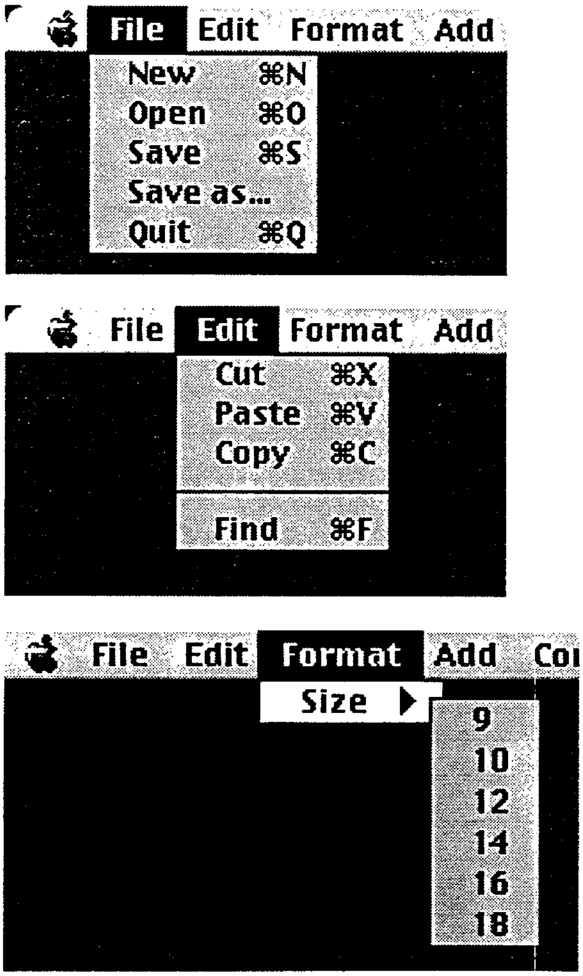


Figure 2. The File, Edit and Format menus

In addition, TransTool contains three special menus for the transcription of spoken language: the Add menu, the Comment menu, and the Tools menu.

The Add menu (Figure 3) contains commands for starting a new utterance (New utterance), for inserting time codes (Time code) and section boundaries (Section), and for marking inaudible speech (Inaudible speech). All of these commands help speed up the transcription process while at the same time minimizing the risk for typing errors and ensuring conformance with the transcription standard.

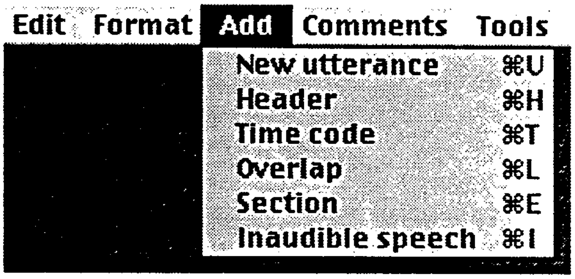


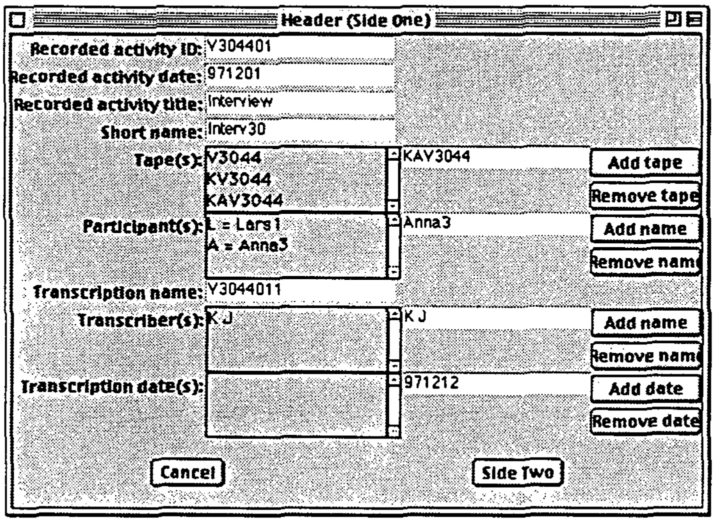
Figure 3. The Add menu

The Add menu also contains a command for marking overlapping speech (Overlap), which automatically inserts and keeps track of the numerical indices used to indicate which portions of speech overlap with each other (cf. Figure 1).

The final command in the Add menu is the command for adding a header to the transcription (Header). This command invokes a set of standardized forms, where the user has to fill in all the relevant information about the recorded activity, such as time and place of recording, type of activity, participants, transcriber, etc. involved in this particular conversation and appriopriate initials for them, transcriber, etc. The forms used to compose the header can be seen in Figures 4 and 5, while the resulting header can be seen in Figure 6.

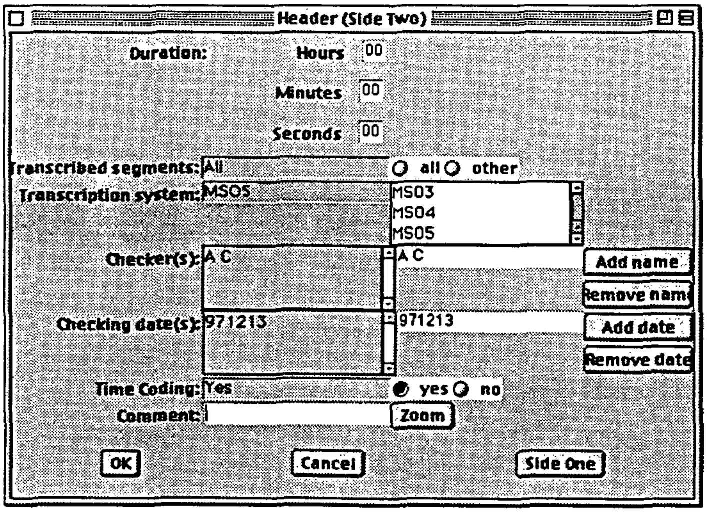
The second special menu is the Comments menu (Figure 7), where the user can select the whole range of standardized comments provided by the transcription standard. The comments are displayed in sub-menus, sorted by category, which may be ripped off and placed as separate windows on the screen. When using this menu, the user first selects the portion of speech that he wants to make a comment about, and then selects the appropriate type of comment from one of the submenus. The comments are automatically indexed.

The final menu of interest is the Tools menu (Figure 8), which mainly contains commands for indexing. In addition to the indexing of comments and overlap (see above), which may need to be updated, the transcription also requires ambiguous spoken language variants (such as the pronunciation 'ja' of the Swedish first person pronoun 'jag') to be disambiguated by numerical indices. This is done through the command MSO indices (where MSO stands for Modified Standard Orthography), which automatically identifies all word forms that need to be indexed and prompts the user for disambiguation.



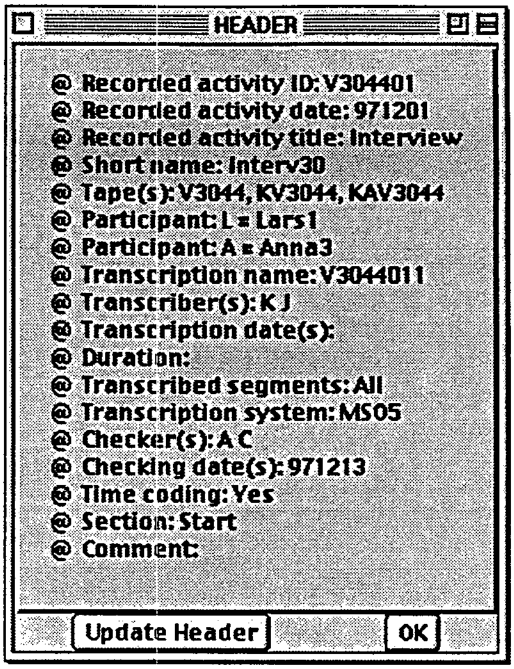
Recorded activity ID: Y304401
Recorded activity date: 971201
Recorded activity title: Interview
Short name: Interv30
Tape(s): V3044
KV3044
KAV3044
KAV3044
Participant(s): L = Lars1
A = Anna3
Anna3
Transcription name: Y3044011
Transcriber(s): K J
K J
Transcription date(s):
971212

Figure 4. Header form (1)



Duration: Hours 00
Minutes 00
Seconds 00
Transcribed segments: All ○ all ○ other
Transcription system: MS05
MS03
MS04
MS05
Checker(s): A C
A C
Checking date(s): 971213
971213
Time Coding: Yes ● yes ○ no
Comment:

Figure 5. Header form (2)



@ Recorded activity ID: V304401
@ Recorded activity date: 971201
@ Recorded activity title: Interview
@ Short name: Interv30
@ Tape(s): V3044, KV3044, KAV3044
@ Participant: L = Lars1
@ Participant: A = Anna3
@ Transcription name: V3044011
@ Transcriber(s): K J
@ Transcription date(s):
@ Duration:
@ Transcribed segments: All
@ Transcription system: MSO5
@ Checker(s): A C
@ Checking date(s): 971213
@ Time coding: Yes
@ Section: Start
@ Comment:

Figure 6. Specified header

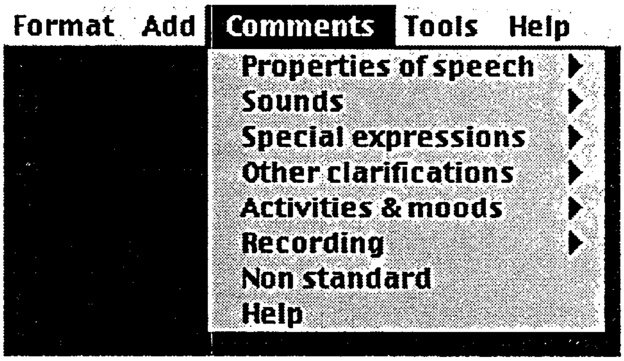


Figure 7. The Comments menu

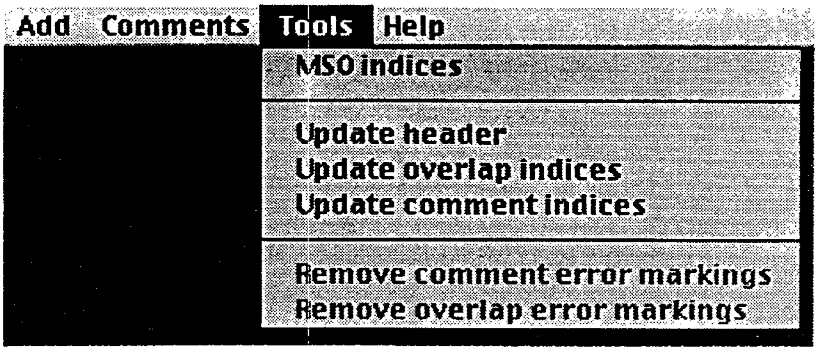


Figure 8. The Tools menu

4. SyncTool

SyncTool is an application developed for synchronizing transcriptions with digitized audio/video recordings. SyncTool is meant to be a synchronizing and viewing tool, allowing the researcher to set time codes in appropriate places in the transcriptions, and to view the transcription and play the recording without having to manually locate the specific passage in the recording.

SyncTool is still in early development, with a limited but working prototype, downloadable from http://www.ling.gu.se/gmslc/. Development is done with cross platform compatibility in mind targeting the Macintosh, Windows and Unix platforms. The prototype has been implemented using a combination of AppleScript and Tcl/Tk on the Macintosh platform; we are currently moving to pure Tcl/Tk and have started a re-implementation in Java.

SyncTool presupposes the following data:

- A transcription conforming to the transcription standard (cf. section 2).
- A media file of some kind, containing the corresponding audio and/or video recording in digitized form.

The user interface is quite straightforward. The user is presented with three windows (Figure 9):

- The Transcript Score & Time Line Window presents the transcription in musical score format along with a time line extracted from the media and media control buttons (bottom window in Figure 9).
- The Media Window (currently an external tool) displays the audio/video recording, allowing the user to swiftly move back and forth in the recording (top right window in Figure 9).
- The Full Transcript Window displays the transcription in original format.

All of these windows, except the Full Transcript Window are available in the prototype we have running. In the final version, while playing an audio or video sequence, the transcription will be scrolled and a visual cue will be shown to indicate which part of the transcription is currently on display. Media controls, such as Play, Stop, etc. will be available, as well as controls for setting the volume, playback speed, stepping back and forth in the recording and looping sequences.

The Speaker Pane inside the Transcript Score & Time Line Window is where the transcription is presented to the user in the special score format used by SyncTool (Figure 10). The score format is a convenient way of displaying an ongoing dialog involving several speakers.

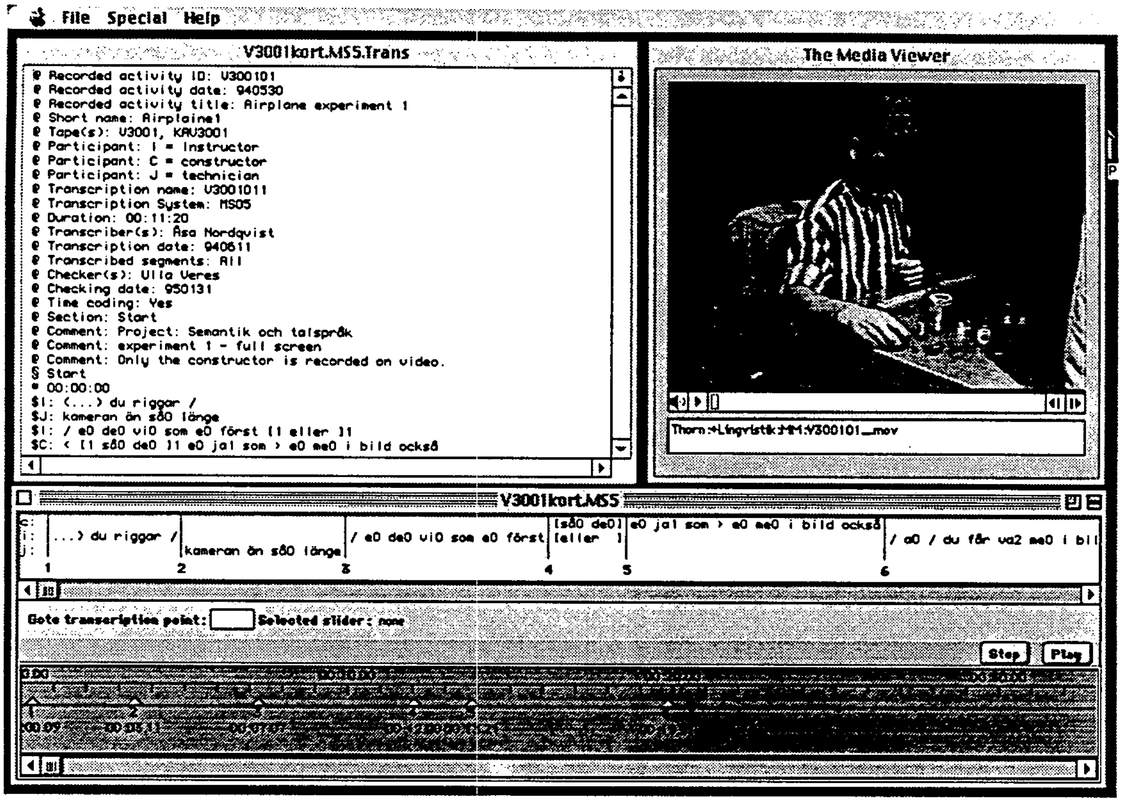
Each speaker is assigned a 'channel' or track, and the utterances that s/he produces are segmented by means of transcription points. Transcription points correspond to speaker changes and the start and end of overlaps. The transcription points used to segement the utterances in the Speaker Pane are derived automatically from the underlying transcription. When a transcription is displayed in musical score format, transcription points span all the speaker channels in the Speaker Pane and are numbered, as can be seen in Figure 11.

The Time Line Pane inside the Transcript Score & Timeline Window allows the user to specify where in the timeline of the recording each of the transcription points occur. Time is measured in minutes, seconds and frames for video with audio, and minutes, seconds and milliseconds for audio only recordings.

A slider for each transcription point sets its time in the recording (as shown by the label beneath it), and can be moved back and forth through the timeline. Transcription point sliders cannot go outside their boundaries, e.g., it is not possible for transcription point 2 to move to the left of transcription point 1, or to the right of transcription point 3, and so on.

In Figure 13 the correspondence between the transcription points in the Speaker Pane and in the Timeline Pane is highlighted with arrows.

Initial placement of sliders is very roughly calculated with a simple algorithm, which distributes the sliders along the timeline more or less according to the length of the utterance. Note that we are not doing any kind of sound signal analysis; the algorithm is based on length of the text appearing in the transcript. The result is, as could be expected, not very good, but it helps somewhat and saves some time; we are working on improving the

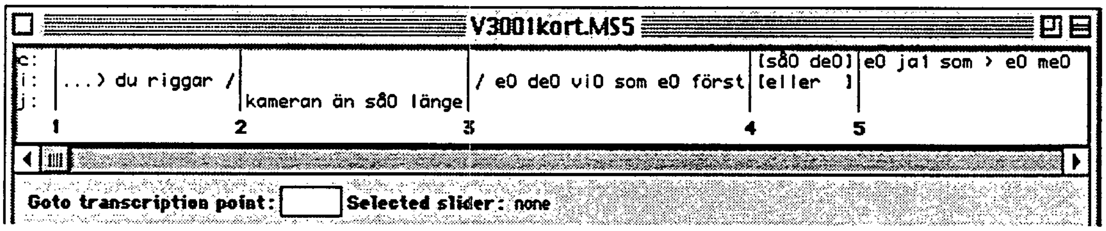


```
@ Recorded activity ID: V300101
@ Recorded activity date: 940530
@ Recorded activity title: Airplane experiment 1
@ Short name: Airplaine1
@ Tape(s): V3001, KAV3001
@ Participant: I = Instructor
@ Participant: C = constructor
@ Participant: J = technician
@ Transcription name: V3001011
@ Transcription System: MS05
@ Duration: 00:11:20
@ Transcriber(s): Åsa Nordqvist
@ Transcription date: 940611
@ Transcribed segments: All
@ Checker(s): Ulla Veres
@ Checking date: 950131
@ Time coding: Yes
@ Section: Start
@ Comment: Project: Semantik och talspråk
@ Comment: experiment 1 - full screen
@ Comment: Only the constructor is recorded on video.
§ Start
* 00:00:00
$I: (...) du riggar /
$J: kameran än såO länge
$I: / eO deO viO som eO först [1 eller ]1
$C: < [1 såO deO ]1 eO jal som > eO meO i bild också
```

```
c:                                                       [såO deO]|eO jal som > eO meO i bild också|
i: |...) du riggar /|                     |/ eO deO viO som eO först|[eller  ]|                          |/ aO / du får va2 meO i bil
j: |                |kameran än såO länge|
    1               2                    3                          4         5                          6
```

Goto transcription point: ☐ Selected slider: none

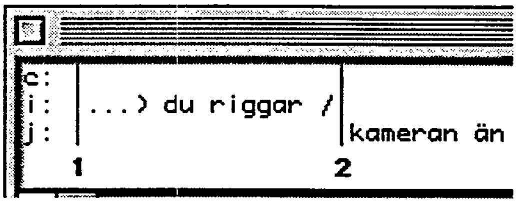
Figure 9. Overview of the SyncTool user interface



```
c:                                                        [såO deO]|eO jal som > eO meO
i: |...) du riggar /|                     |/ eO deO viO som eO först|[eller  ]|
j: |                |kameran än såO länge|
    1               2                    3                          4         5
```

Goto transcription point: ☐ Selected slider: none

Figure 10. Speaker Pane, transcription in score format



```
c:
i: |...) du riggar /|
j: |                |kameran än
    1               2
```

Figure 11. Channels in the Speaker Pane; three speakers: c, i and j

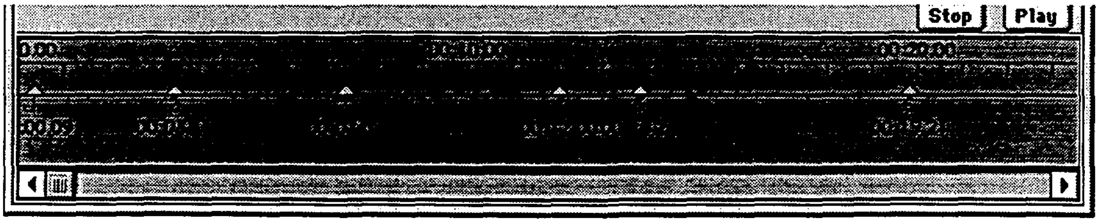


Figure 12. Timeline Pane, with transcription point sliders

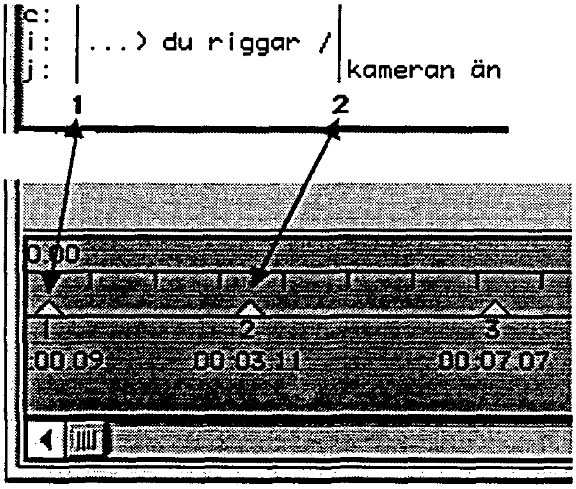


Figure 13. Timeline Pane, with transcription point sliders

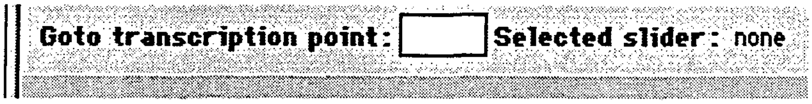

Goto transcription point: Selected slider: none

Figure 14. Go to transcription point

algorithm, possibly using words per second measures and/or involving simple sound signal analysis.

To go to a specific transcription point, the user enters it in the 'Go to transcription point' edit field (Figure 14). Both the Speaker Pane and the Timeline Pane then scroll to the transcription point in question.

Double clicking on a slider plays the recording from that transcription point. To stop playback, the user presses the space key on the keyboard. The prototype also implements primitive playback controls, currently only Start and Stop. These will be enhanced in upcoming versions.

The Media Window is currently implemented as an external application, The Media Viewer. It has no user interface apart from the media controllers, and its only purpose is to allow other applications to communicate with it and programmatically tell it to open media files, start/stop playback, provide data on the movie, etc. It uses QuickTime and all the media formats that QuickTime supports (i.e. MPEG, QuickTime movies, AIFF, WAV, etc.), and it provides full motion playback of

MPEG-1 movies with appropriate computer hardware.

The Media Viewer is available as a separate Macintosh application. Note however that we are in the process of implementing the functionality of The Media Viewer into SyncTool as plug-in module, which we hope will give us better control of how media files are used in SyncTool.

Planned, but currently missing features include the possibility to visualize sound waves along the Time Line Pane, which will let users more easily position the transcription point sliders. We are also considering the inclusion of spectrograms, etc. Furthermore, we need to implement the real-time visual cue indicating which part of the transcription is being played back. This will be much easier to achieve when the Media Viewer functionality is built into SyncTool. Another feature that is going to be implemented is the possibility to add and delete transcription points, as well as separate tracks for different kinds of synchronisations (gestures, etc).

In its current state, SyncTool is not only a synchronising tool, it is also a viewer of synchronized transcriptions. The tool presents audio, video and text simultaneously. The user can select parts of the transcription text and have the corresponding audio or video sequence played back with a minimum of effort. With this use in mind, the possibility to correct errors found in the transcription, such as incorrectly marked overlaps, will be a very useful feature too. The prototype has already shown its usefulness in this area, highlighting the usability of the score format. These errors are quite hard to discover using the traditional full view of a transcription. Therefore, it is possible that TransTool and SyncTool will be merged in the future. Ultimately, our goal is to provide a set of integrated tools for transcription, synchronization/alignment, coding, annotation and presentation.

5. Some Comparisons

In this section, we will compare our tools to some other programs that help align text and recordings. The tools considered are:

- SoundWalker (by Jack Du Bois) [http://humanitas.ucsb.edu/depts/linguistics/research/csae/soundwalker/walk.html]
- SoundWriter (by Jack Du Bois) [http://humanitas.ucsb.edu/depts/linguistics/lab/transcriptions.html]
- SyncWriter (by med-i-bit) [http://www.med.i.bit/Software/syncWRITER/info.english.html]

SoundWalker/VoiceWalker/MediaWalker lets you view recordings in 'auto-pilot' mode; we quote from the on-line manual: 'The most distinctive feature in SoundWalker for controlling the playback of recorded speech is called the Walk. The Walk function plays the recording in manageable chunks so that the transcriber can concentrate on transcribing, as it automatically Walks through the recording one Step at a time. It plays a brief sound bite consisting of the first four seconds of the recording (one Step), and repeats this portion of sound several times to allow the user to transcribe it. Then it steps forward slightly, beginning the second Step about one second after the first. It plays this new four-second chunk of sound several times, and then moves on to the third Step. Because each new Step overlaps partially with the previous one, the transcriber always has enough familiar context to know where s/he is in the recording. And because the Walk is entirely automatic, it leaves the user's hands free to transcribe using his/her preferred word processor in a separate window.' The user interface of SoundWalker is depicted in Figure 16.

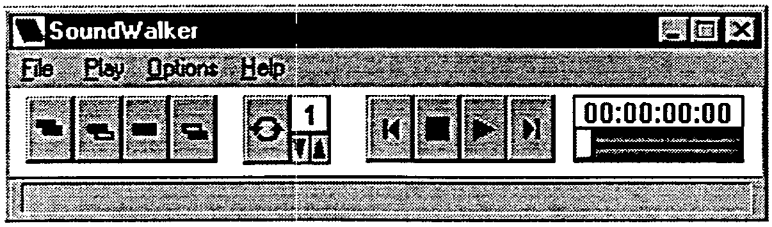


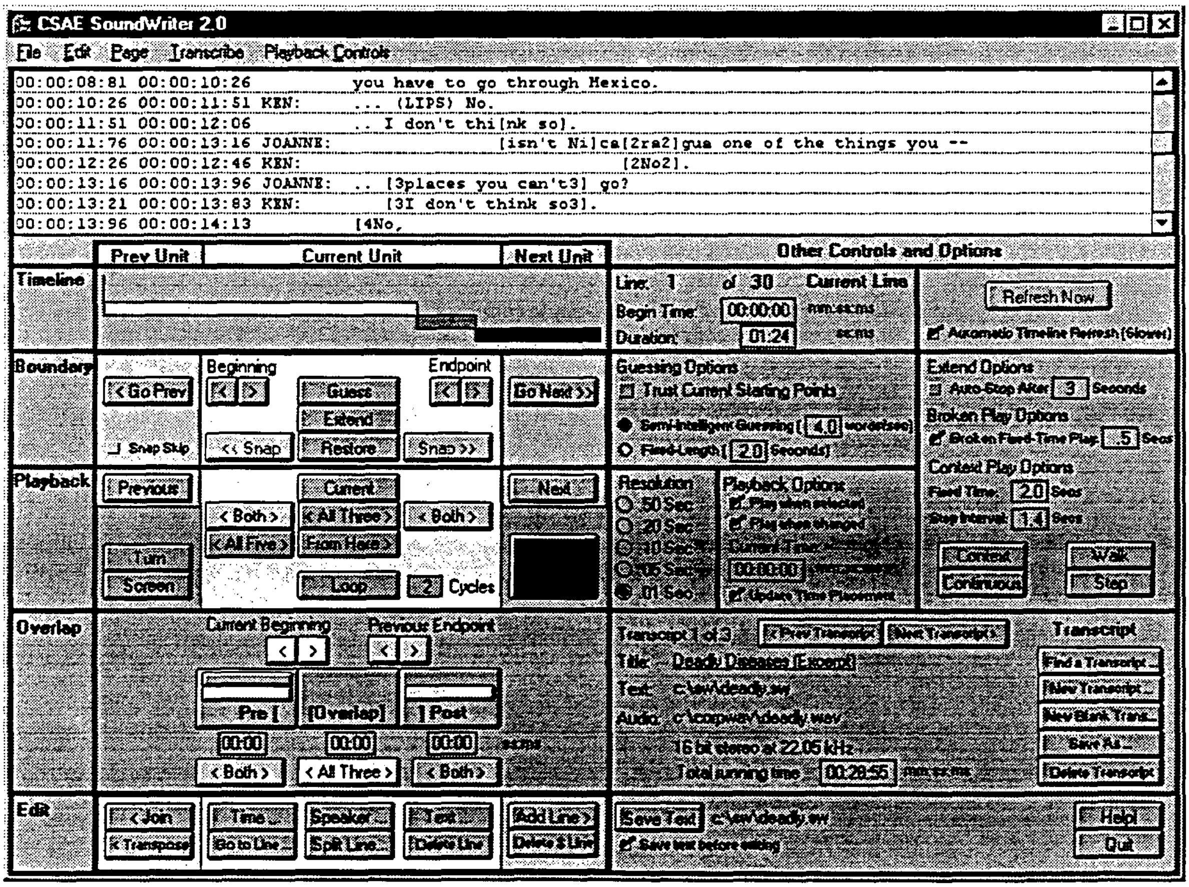
Figure 16. The SoundWalker/MediaWalker user interface

Compared to our tools, SoundWalker provides a subset of the functionality that we plan to provide in SyncTool, albeit in a more refined and elegant way, which we have not achieved yet. The main focus of SoundWalker is to support the manual transcription process and as such that functionality should be provided in Transtool.

As was mentioned above, we are already planning to integrate TransTool and SyncTool The development of The Media Viewer into a plug-in module for inclusion in Transtool is one step towards that goal. SoundWalker uses Word as its text processor, something that we cannot do. When transcribing spoken language we use our own Modified Standard Orthography (MSO), which lets us transcribe speech as it is actually pronounced. MSO then uses an indexing system to map between Standard Orthography and MSO. This indexing feature, along with automated overlap handling, is managed by TransTool.

Turning from SoundWalker to SoundWriter, we first note that: 'SoundWriter incorporates the features of SoundWalker 1.1 as well as the ability to align transcripts with sound files. Basically, this program assigns starting and ending SMPTE time codes to each intonation unit.' (From download page.)

SoundWriter provides more or less the basic functionality we want to have in SyncTool and in addition allows the user to partially edit the transcription. The alignment tool of SoundWriter is very nice, and we do not have anything like it in SyncTool or The Media Viewer today. Something that is particularly helpful is the 'guessing' function in SoundWriter. Even if it is not a guessing function per se (you specify the number of words per second, and then it 'guesses' where the next turn is) it clearly speeds up the alignment of transcription and recording.



```
00:00:08:81 00:00:10:26          you have to go through Mexico.
00:00:10:26 00:00:11:51 KEN:     ... (LIPS) No.
00:00:11:51 00:00:12:06          .. I don't thi[nk so].
00:00:11:76 00:00:13:16 JOANNE:                 [isn't Ni]ca[2ra2]gua one of the things you --
00:00:12:26 00:00:12:46 KEN:                                [2No2].
00:00:13:16 00:00:13:96 JOANNE:  .. [3places you can't3] go?
00:00:13:21 00:00:13:83 KEN:        [3I don't think so3].
00:00:13:96 00:00:14:13          [4No,
```

Line: 1 of 30 Current Line
Begin Time: 00:00:00 mm:ss:ms
Duration: 01:24 ss:ms

Guessing Options: ● Semi-Intelligent Guessing [4.0] words/sec; ○ Fixed-Length [2.0] Seconds
Extend Options: Auto-Stop After [3] Seconds
Broken Play Options: Broken Fixed-Time Play [.5] Secs
Context Play Options: Fixed Time: [2.0] Secs; Stop Interval: [1.4] Secs
Resolution: ● .01 Sec

Transcript 1 of 3
Title: Deadly Diseases (Excerpt)
Text: c:\sw\deadly.sw
Audio: c:\corpwav\deadly.wav
16 bit stereo at 22.05 kHz
Total running time 00:28:55 mm:ss:ms

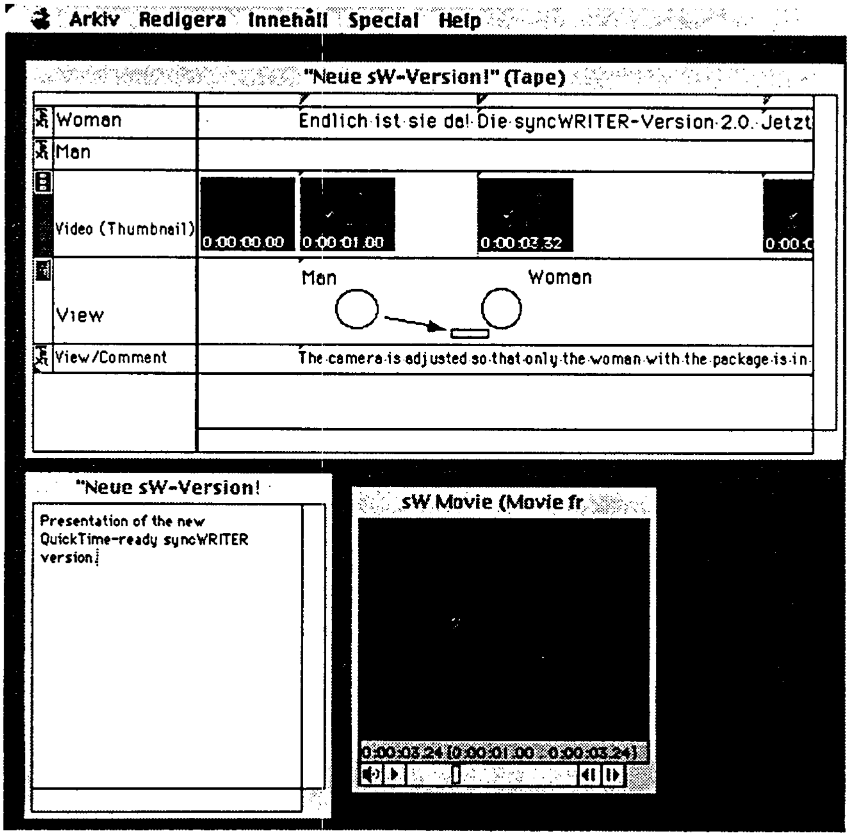
Save Text c:\sw\deadly.sw
☑ Save text before exiting

Figure 17. The SoundWriter user interface

There are also differences, however. The most important one is in the transcription standard used. From the information that can be gathered from the web site, it seems that this standard only covers a small subset of the phenomena taken into account by our standard. Another difference is the musical score format used in SyncTool but not in SoundWriter. Moreover, SoundWriter does not support the use of video recordings in the same way as SyncTool does. Finally, SoundWriter is not platform independent but only exists for Windows computers.

The third program, SyncWriter, handles texts with simultaneous passages (tracks or channels) and works with the notion of a musical score format in a way similar to SyncTool. Figure 18 shows a screenshot of the SyncWriter user interface layout.

SyncWriter does what we need to do; it synchronizes text with a QuickTime movie. There is a Tape window (the topmost window in Figure 18) that contains all the text tracks, the movie track(s) and whatever extra tracks one deems necessary. It is possible to attach a movie to the movie track. A thumbnail of the movie is then displayed in the track.



Endlich ist sie da! Die syncWRITER-Version 2.0. Jetzt

The camera is adjusted so that only the woman with the package is in

Presentation of the new QuickTime-ready syncWRITER version.

Figure 18. The SyncWriter user interface

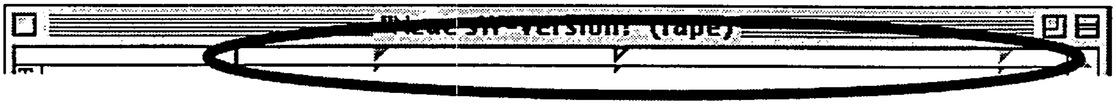


Figure 19. Synctabs

One of the drawbacks with SyncWriter is that you have to synchronize all of the tracks separately, as there is no hierarchy of tracks or of synctabs. This can be very time-consuming if you have a lot of speakers. Moreover, there is no time line, you attach a movie to a movie track. And the system with movie thumbnails is not very elegant.

So, even if SyncWriter contains some of the features that we want TransTool and SyncTool to have, it is not quite adequate for our needs. Finally, it is again platform specific (Macintosh).

6. Conclusion

In this paper, we have argued for the usefulness of an integrated platform for multimodal spoken language corpora, and we have presented two simple tools that have been developed as components of such a framework. Although these tools are still far from constituting a full-fledged platform for multimodal spoken language corpora, with synchronized display of transcriptions and audio/video recordings, as well as tools for annotation and presentation, they nevertheless represent the first steps towards such a platform and have already proven useful in their own right. We also believe that the experience gained from the development of these tools will be valuable in future work towards a more ambitious and useful toolbox.

References

Allwood, J. (1998) Some Frequency-Based Differences between Spoken and Written Swedish. To appear in Proceedings of the XVIth Scandinavian Conference of Linguistics, Turku, November 1996.

Nivre, J. (1998) Transcription Standard. Version 5.2. Technical Report. Göteborg University: Department of Linguistics.

SoundWalker:
http://humanitas.ucsb.edu/depts/linguistics/research/csae/soundwalker/walk.html

SoundWriter:
http://humanitas.ucsb.edu/depts/linguistics/lab/transcriptions.html

SyncWriter:
http://www.med.i.bit/Software/syncWRITER/info.english.html