ISSN 1974-4382

METHODOLOGICAL CHALLENGES IN AUDIOVISUAL TRANSLATION: EXPERIMENTING NEW SOFTWARE FOR MULTIMODAL CORPUS-BASED ANALYSIS

DORA RENNA
CA' FOSCARI UNIVERSITY OF VENICE

dora.renna@unive.it

Citation: Renna, Dora (2022) "Methodological Challenges in Audiovisual Translation: Experimenting New Software for Multimodal Corpus-based Analysis", in Lucia Abbamonte and Flavia Cavaliere (eds.) *Updating Discourse/s on Method/s*, *mediAzioni* 34: A199-A219, https://doi.org/10.6092/issn.1974-4382/15557, ISSN 1974-4382.

Abstract: Audiovisual translation has long struggled to strike a balance between corpus-based analysis of large amounts of text and the need to systematically integrate multimodality in its research scope, in order to fully acknowledge the complex nature of the audiovisual product. This paper aims to relate on the experimentation that is currently being conducted at Ca' Foscari University of Venice, with the collaboration of the University of Basel (Switzerland): using an existing software (created for pragmatics) for audiovisual translation. The main aim is to make sure this software is able to support the researcher in transcribing, annotating, adding metadata, managing and querying text and video files. The experimented software has never been used for parallel aligned audiovisual text so far. The contribution here presented first briefly describes a framework developed to analyse language variation and multimodality (with a focus on character design) in audiovisual translation, which was then immersed in the software. Subsequently, the software itself is described in detail, with specific attention to its potential and limits in the use within the field of Translation Studies and audiovisual translation. This is done by showing examples from a pilot study that belongs to a broader corpus currently under construction.

Keywords: audiovisual translation; digital humanities; multimodality; linguistic variation.

1. Introduction

By their own nature, audiovisual products "involve the simultaneous deployment of several sign repertoires (including but not limited to speech or the written language of subtitles, film editing, image, music, colour or perspective)", perceived by the audience simultaneously through "the synchronized use of multiple media" (Pérez-González 2014: 186-187). This is the essence of multimodality, intended both as "the combination of speaking, writing, visualization and music" (Pérez-González 2014: 185) and the investigation of these modes and the way(s) they combine (Bateman 2014: 6). Multimodality is the distinguishing feature of audiovisual texts, whose meaning goes beyond the sum of its parts, since each part can complete, explain, multiply, and modulate the others (Lemke 1998: 92). Here lies the fascination of audiovisual products, as well as their complexity. There is a stark contradiction between the naturally holistic fruition of such products and the difficulties of dealing with them when it comes to analysis – which, by definition, requires to set things apart in order to understand them. The result is that "'texts' and 'images' are generally addressed by very different groups of disciplines and those disciplines often take different views both on just what they are describing and how descriptions should be pursued" (Bateman 2014: 5). Translation Studies (TS) and in particular the field of Audiovisual Translation (AVT) have long suffered from this separation (Chaume 2004; Gambier 2006; Díaz-Cintas 2008). While multimodality can successfully be integrated in single case studies, the attempt to do so with a descriptive, corpus-based approach seems to pose even greater challenges (Ramos Pinto and Mubaraki 2020). The reason should probably be found in the mainly text-focused tools that are available to researchers – bearing in mind that tools like media, languages and technology work as extensions of human mind and culture (McLuhan 1964).

Multimodality becomes perhaps even more central when translation scholars and practitioners are faced with the delicate issue of on-screen representation of minority groups in films. Here, in fact, language and non-verbal aspects of an audiovisual product work in synergy and can become very influential, given the power of cinema to shape the collective memory (Fluck 2003: 224). The issue of mediated *images* of minority groups within a society has been dealt with in various fields, including gender studies and media studies (e.g. *controlling images*, Hill Collins 2002; *poetics and semiotics of stereotyping*, Ramírez Berg 2002) and has recently been acknowledged by TS with the study of *imagology* (van Doorslaer et al. 2016). Like the name of the discipline itself suggests, this type of study would immensely benefit from a multimodal approach. The need to integrate qualitative single-case studies with broader corpus-based studies lies in the descriptiveness that is typical of corpora. Descriptiveness in TS helps overcome the assessment perspective, which is certainly useful but may, in some cases, hinder understanding and contextualisation (Laviosa 2002: 14).

That is why this article sets out to contribute to a shift in AVT studies by relating on the experimentation currently being conducted at Ca' Foscari University of Venice: using the software EXMARaLDA (Extensible Markup Language for Discourse Annotation) for audiovisual translation. The software,

https://doi.org/10.6092/issn.1974-4382/15557

born for analysing oral corpora and broadly used in pragmatics, had never been experimented in the field of AVT. Such an effort was possible thanks to the precious collaboration and training provided by one of the main software developers, Thomas Schmidt (University of Basel, Switzerland), with the support of the Department of Linguistics and Comparative Cultural Studies (Ca' Foscari University of Venice). Section 2 provides an outline of a framework designed to analyse language variation and character design in AVT with a multimodal, corpus-based approach. Most space is dedicated to Section 3, where the main features of the software, their relevance to the framework at hand and to TS scholars in general are explored in detail. Finally, some conclusions concerning methodology are briefly collected in the last section.

2. A framework to analyse language variation and character design with a corpus-based multimodal approach in AVT

According to Ramos Pinto and Mubaraki (2020), the focus of AVT on the written mode is especially apparent in corpus-based research (today particularly appreciated by TS scholars). Corpora techniques (and technologies), however, have mainly focused on textual analysis. The authors note that audiovisual translation studies tend to focus either exclusively on text in order to analyse large corpora, or to use qualitative (and smaller) case studies to look at a broader range of aspects of the audiovisual product. The construction and even more the processing of a multimodal corpus is a particularly demanding task, as the objective should not be to just store the audiovisual material, but also to make it analysable and processable (Allwood 2008: 208). In this sense, Iedema (2003: 48) stresses a key point in multimodal studies, namely that the focus should be "the relationships between […] different semiotics, and on the "division of labour" between them in particular representations". This alludes to the importance of building a bridge between different modes, using relationships as linchpins. Ramos Pinto (2018) and Ramos Pinto and Mubaraki (2020) elaborated this concept, also informed by the multimodal work of Pastra (2008). Their breakthrough work is a key reference in the development of the framework presented here.

The model consists of three dimensions (Ramos Pinto 2018; Ramos Pinto and Mubaraki 2020; Renna 2021). The first is the textual dimension, focused on linguistic *variation* and its *realisation*. The second is the diegetic dimension, focusing on the *intermodal relations* established between textual dimension and a selection of cinematic elements that are considered relevant for image construction. The third dimension is the sociocultural one, allowing to interpret both source text (ST) and target text (TT) in the light of the *context* that generated them. Since the first two dimensions are the most quantitative and product-based,[1] they were at the core of the software experimentation and will be explored here.

[1] The third one is conceived as the final step of the analysis, and will involve perception studies aimed at verifying the product-based findings.

https://doi.org/10.6092/issn.1974-4382/15557

The aim is to show how linguistic variation and multimodality interact in the ST to outline the mediated image of a minority, or *ethnotype* (van Doorslaer et al. 2016),[2] and how such interactions are (re)presented in the TT. While the first experimentation was carried out on a synchronic case study with manual tagging and two different data processing software for statistic elaboration (Renna 2021), the current experimentation has been conducted on a pilot study that is part of a larger research project at Ca' Foscari University of Venice. The project is a diachronic study of the cinematic image of characters of Chinese[3] origin in Hollywood films and Italian dubbing.[4] In particular, the pilot study was the film *Charlie Chan in London* (1934) and its Italian dubbed version, *Il nemico invisibile*, as this was the oldest film in the corpus and had some crucial features for the analysis. The protagonist is Charlie Chan, a character notoriously speaking a strongly stereotyped variety of English and using an array of stereotyped mannerisms (Huang 2010; Ma 2014). Moreover, since the film is rather short (less than 80 minutes) and Chan is the only character of Chinese origin in the film, the relatively limited number of minutes to analyse seemed ideal for a pilot.

The framework is designed to answer the following questions: 1) what fictional variety does the character speak in ST and TT? 2) which marked features characterise this variety? 3) which strategies were adopted in dubbing? 4) how do intermodal relations contribute to meaning making in ST and TT? Since the study focuses on character design, only the relevant character is sampled out of the whole film. The unit of analysis chosen for dubbing is the line, spotted according to cinematic conventions such as turn taking (also signalled though body language), long interruption (usually >3 seconds) with new informative material being introduced, or scene change (Gurskis 2006: 180).

Once the text has been transcribed and divided into lines, the textual dimension analysis requires to attribute each line to a *fictional* linguistic variety. It is important to underline that in fiction the language is not a realistic representation of spontaneous speech, but rather "the creative use of linguistic varieties" used as "a textual resource that helps the reader to define the sociocultural outline of the character, as well as his/her position in the sociocultural fictional context" (Ramos Pinto 2009: 291). This implies that real-life taxonomies might not be satisfactory when dealing with fiction, and that is why the classification applied to this pilot is designed for fictional variation (Renna 2021) and based on two main criteria: 1) belonging to a group considered

[2] The study does not focus on the diasporic identity of the characters, but on their mediated *images* or *ethnotypes*, intended as "rationalizations of cultural difference" (van Doorslaer et al. 2016: 3).

[3] Since, as previously stated, the analysis focuses not on identity but on ethnotypes, it might be important to define here some key concepts concerning terminology. The terms "Chinese American" and "Chinese diaspora" are referred to overseas Chinese (*Huáqiáo*, 华侨), living in the United States, as well as to American citizens of Chinese origin. The word "Chinese" refers here to all ethnic Chinese. While a deep debate on identity is certainly important, it is not the aim of this methodological article. For further reding on the matter, some crucial readings (that I used as a reference in my studies) are Lisa Lowe (1991), Yến Lê Espiritu (1992), or Peter Feng (2002).

[4] The project title is *A Multimodal Study on the Construction and Perception of Linguistic Variation and Stereotypes in Audiovisual Translation*, and it is being carried out at the Department of Linguistics and Comparative Cultural Studies of Ca' Foscari University of Venice under the supervision of Francesca Santulli, Linda Rossato, and Elisa Bordin.

https://doi.org/10.6092/issn.1974-4382/15557

either as *dominant* or *minority* within the diegetic context considered (in the pilot study, White Anglo-Saxon Protestants or people of Chinese origin respectively); 2) adherence to what is considered as the "standard" variety in that fictional world. In this case study, the ST standard is the mid-Atlantic accent, particularly appreciated in 1930s cinema (Naremore 1988: 49). The standard in the Italian TT is *dubbese*, a fictional variety closer to written and not used by speakers outside of dubbing (Antonini 2008: 136). These two criteria define the prestige of a variety: the more a character occupies a central position in terms of language, the more prestigious their sociocultural positioning (Ramos Pinto and Mubaraki 2020). The scheme used for the classification is the following:

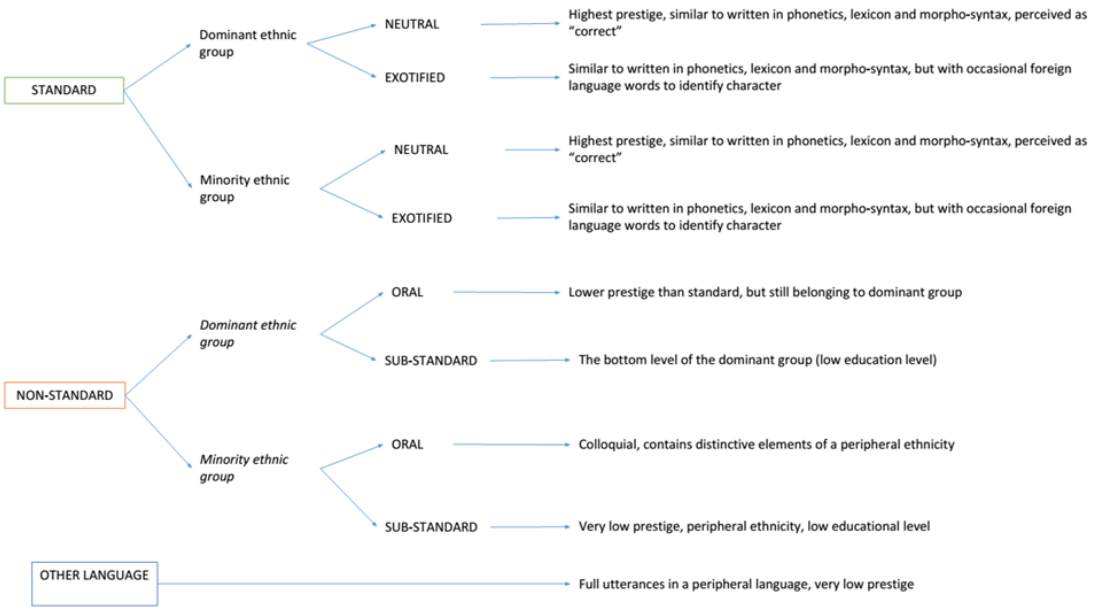


Figure 1. Linguistic variation in films (Renna 2021)

Each line can only be attributed to one option (*other language* is the option for code switching to heritage language, not used in this pilot). Subsequently, the marked elements are identified – more options can be selected – phonetics (and suprasegmental traits), morphosyntax and/or lexicon. The query phase with EXMARaLDA suggested adding other linguistic categories: 1) discourse, intended as verbal communication features attributable to an ethnotype; 2) proverbiality (utterances recalling chéngyǔ成语 and proverbs, be it real Chinese sayings or made-up ones); 3) qǐng-like calques from Mandarin (请).

The diegetic dimension sheds light on the intermodal relations between textual dimension and the broader audiovisual context. Conceiving multimodality in terms of relation (rather than, for example, description) allows a straightforward data elaboration, easily manageable by individual scholars or small groups (Ramos Pinto and Mubaraki 2020). However, while a film is a complex unit of different elements perceived and processed simultaneously by the audience, to analyse means to isolate the most relevant parts of the audiovisual whole. Drawing from by Ramírez Berg's *poetics of stereotyping* (2002), the selected elements are: 1) costumes and make up, the latter also comprising scene make up

https://doi.org/10.6092/issn.1974-4382/15557

(e.g. dark circles for a tired look); 2) phenotype (either real or ostensible) that hints at the belonging to a certain group; 3) behaviour, the way a character marks their presence on the scene; 4) interlocutors, i.e. the other characters participating in the interaction; 5) setting, including a series of cinematic features like *mise-en-scene*, framing and lighting; 6) music, both intended as soundtrack and music played in the scene. During the query phase, a need to focus on the character's proxemic behaviours emerged, and a category regarding "proxemics" was added to the analysis. The range of possible relations is drawn from Ramos Pinto and Mubaraki's (2020) adaptation of Pastra's (2008) scheme to AVT studies. The main ones are confirmation and contradiction, based on whether the linguistic variety used is expected within the diegetic context. Confirmation can be further divided between equivalence or complementarity, depending on whether the verbal and non-verbal aspects overlap or complete each other. For each line only one option is chosen for each of the non-verbal elements.

3. A multimodal software for a multimodal analysis: using EXMARaLDA in AVT

EXMARaLDA (Extensible Markup Language for Discourse Annotation) was chosen for the experimentation whose aim was to join the various aspects of the complex framework here presented in one software environment, while keeping both language variation and multimodality at the core of my AVT studies. EXMARaLDA is a software with a strong multimodal vocation, which allows working with corpora of spoken language. It was conceived with multiple aims. First, to streamline the exchange of spoken language corpora among different technological environments; second, to use multimedia and hypertext in language analysis; finally, to host large and reusable corpora (Schmidt and Wörner 2009: 565-566).

The software includes three main tools,[5] each dedicated to a specific phase of corpus-based analysis:

- Partitur Editor, the tool for transcription,[6] editing, annotation and output of multimodal corpora;
- Corpus Manager (CoMa), for corpus management and organisation, metadata entry, corpus bundling and subdivision;
- Exakt, the corpus analysis and concordance tool, which allows to query and quantify data, as well as to export results.

These three tools are interdependent, as they work like phases of an assembly line, where each delivers a semifinished product to be further elaborated inside the next. Developed starting from 2000 at the Research Centre for Multilingualism in Hamburg, EXMARaLDA was used for a series of spoken corpora. The software website[7] contains downloads for different operating systems (regularly updated),

[5] There are also other tools: FOLKER, a simplified version of Partitur Editor optimised for spoken German, and OrthoNormal, for normalisation of FOLKER transcripts. Since they are especially dedicated to German, they will not be analysed in detail in this article.
[6] EXMARaLDA is suitable for different transcription conventions and features a keyboard for special characters.
[7] Available in English and German: https://exmaralda.org/ (visited 19/07/2021).

https://doi.org/10.6092/issn.1974-4382/15557

guides and video tutorials (in English or German), as well as a list of publications about EXMARaLDA and its applications. It also contains an email address to contact the developers for any queries.

Probably the previous most relevant experiences to the present study are the EXMARaLDA-based corpora in the fields of pragmatics, bilingualism, interpreting and written-text translations (Hedeland et al. 2011: 230-232). While each of these corpora had something in common with the corpus at hand, the software had never been tested with AVT studies before. There are programs that can be compared to EXMARaLDA, and that were also taken into consideration for this corpus. In particular, ELAN (designed for sign language and endangered languages) and PRAAT (specialised in phonetics) have similar functions.[8] The software here presented can communicate with both, as it is possible to import whole ELAN an PRAAT corpora into EXMARaLDA and vice versa. One of the strongest advantages of EXMARaLDA, however, lays in the three-tool structure, which makes each function particularly powerful and suitable for larger corpora.

While one may argue that, from a scholarly point of view, audiovisual translation and pragmatics (the main subject treated with EXMARaLDA) are not necessarily close fields, from a software perspective there are some common traits. One is the interest of pragmatics in "linguistic behaviour on *different linguistic levels*" (Schmidt and Wörner 2009: 567). The other is the centrality of context to infer meaning: Schmidt and Wörner (2009: 567) distinguish among *interactional context* (provided by behavioural data), *situational context* (the more general circumstances in which the event takes place) and *ethno-graphic metadata* (background information about the speakers and their social relationships). A film has all these features, the main difference being its fictional nature. Fictionality, however, does not make this information any less important: indeed, as can be evinced from the previous section, they are purposefully laid in the scene to integrate the communicative meaning of the diegesis.

It might be worth looking at the tools one by one to explore the way they can be useful for multimodal and corpus-based audiovisual translation studies. Each tool is first described, with specific attention to the features that were relevant for AVT. Subsequently, the outputs and results obtained in each phase are outlined.

3.1. AVT transcription and annotation with Partitur Editor

3.1.1. Description and uses for AVT studies

As a transcription tool, Partitur Editor is based on a musical score outlook, "which best meets requirements for representing multi-party, multi-level and multi-modal descriptions of spontaneous interaction" (Schmidt and Wörner 2009: 569) or, like often happens in audiovisual translation, of fictional interaction that is mostly aimed at seeming spontaneous. It is possible to import transcripts into Partitur Editor, provided they are plain text files (.txt).

[8] More information on ELAN and PRAAT can be found on their respective webpages, https://archive.mpi.nl/tla/elan (visited 19/07/2021) and https://www.fon.hum.uva.nl/praat/ (visited 19/07/2021).

https://doi.org/10.6092/issn.1974-4382/15557

In the Partitur Editor transcription, different layers extend both horizontally (in time progression from left to right) and vertically, representing simultaneous events and/or different analytical aspects of the same event. Each line, called *tier*, can be assigned a series of functions, the most important being transcription, description, and annotation. Tiers are attributed to a *speaker*. The segment of speech that is transcribed and annotated in a time section is called an *event*. Transcription and annotation are attached to the specific moment they are referred to, both in terms of time and of position in the video clip, through a similar process to the one of spotting and input in subtitling. Partitur Editor features a sophisticated video player[9] that allows fine-grain spotting and proves to be particularly usable throughout transcription and annotation. In addition, more than one video and audio track can be added to the same transcription file, which is particularly important for AVT scholars, who usually deal with at least two versions of the same audiovisual product.

With regards to my framework, each Partitur Editor file only contains the transcription of a film sample, represented by one of the characters selected for the analysis. In the pilot study only one character (Charlie Chan) was analysed. However, in the other films currently being analysed, more than one character per film have been selected. Since the focus of my research is character design, each of them (in both ST and TT version) will be featured in a different transcription file. This will also allow the file to remain relatively manageable, given the considerable number of annotation tiers.

Based on the experience with the pilot study and the dialogue with the system developer, a good way to achieve alignment is to have ST and TT version of the same character as two speakers superimposed in the same file, with the relative annotation tiers following each (see red arrows in Figure 2).

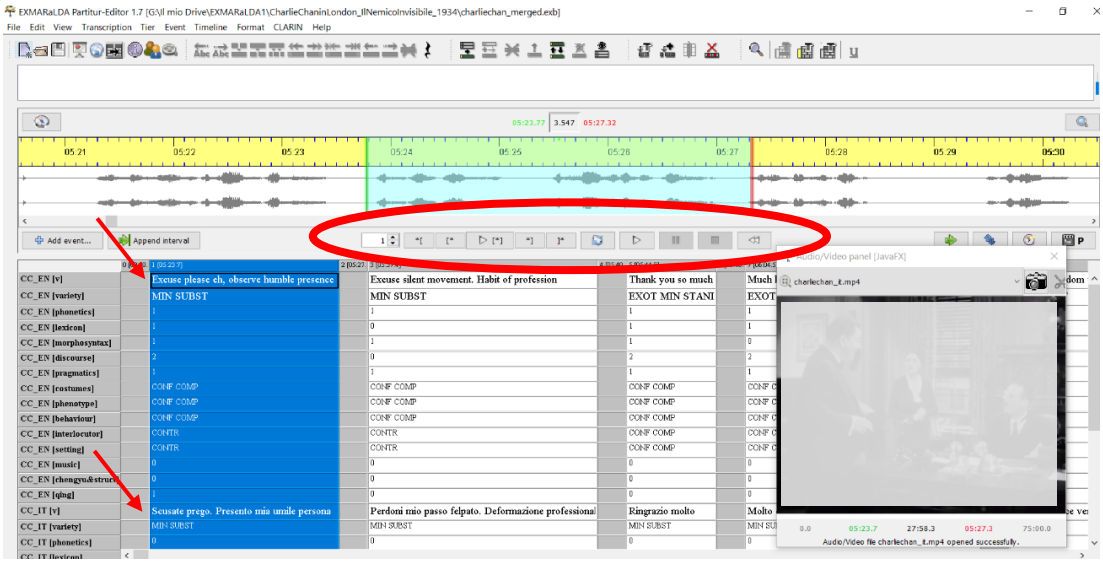


Figure 2. A capture from the pilot study in Partitur Editor. Media player options circled in red, ST and TT version as two aligned speakers marked by red arrows.

[9] The possibility to rewatch the specific scene linked to an event will also come back in Exakt, keeping the corpus multimodal throughout the process.

https://doi.org/10.6092/issn.1974-4382/15557

EXMARaLDA, in fact, is also designed for the transcription of overlapping speakers (be it in the same video and in two different videos). While this may not be the ideal approach for other types of study, it proved particularly useful in analysing linguistic and multimodal character design.

In the capture[10] from the pilot in Partitur Editor (Figure 2), it is possible to visualise the way EXMARaLDA has brought to life the framework presented in the previous section. To the left, in grey, are the categories (which remain visible when moving forward in the transcript). CC_EN and CC_IT are the ST and TT versions of the character Charlie Chan (CC) respectively. The letter [v] represents the transcription tier, while the others are the levels of analysis of textual and diegetic dimension. The annotation levels are located right after the transcription. I used different closed textual options to tag the linguistic variation and intermodal relation categories (here, MIN SUBST stays for minority substandard, see Figure 1, while CONF COMP and CONTR for confirmation-complementarity and contradiction respectively). In the other cases I used numbers (from 0 to 3) that signalled the absence (0), presence (1) or peculiar features (2 and 3) of said categories. A file with a complete key to read the annotated transcription is added as metadata to the CoMa file. Using strictly codified categories rather than lengthy descriptions allowed speeding up the annotation process – the descriptions are indeed unnecessary since the software allows watching the line segment or the entire scene.

3.1.2 Outputs and results

Partitur Editor's output is the multimodal transcription (extension: .exb, file type: EXMARaLDA Basic Transcription). Even before reaching the following steps (CoMa and Exakt), it is possible to output the transcription in different formats, including HTML (and HTML5 for video clips) for web browsers, RTF for MS Word files, PDF for printing (Schmidt and Wörner 2009, 570). The transcription can also be output in the form of a list by saving it as a XLS stylesheet (Schmidt and Wörner 2009: 571). The transcript can be exported in similar software like ELAN or PRAAT or saved as a subtitle file. Partitur Editor is also able to draw a word list based on the transcription, useful to check for errors but also as a first tool of analysis that may help identify phenomena that will be explored in detail in the following phases. However, as will be explained in the following sections, the main use imagined for the transcriptions is to be organised using CoMa, the corpus management tool, and then queried using Exakt.

Overall, the use of Partitur Editor had some important advantages that can really make the difference for an AVT scholar's workflow. The most evident is the multimodal tagging, simple but sophisticated, which will be particularly comfortable for those who have practiced subtitling. From spotting to transcription and tagging, the software really supports this potentially tiresome work, especially thanks to the media player. Rich in

[10] The movie stills from the pilot case study *Charlie Chan in London* (1934) are covered for copyright reasons.

https://doi.org/10.6092/issn.1974-4382/15557

options (see area circled in red in Figure 2) such as loop player, section player, and even buttons that allow playing the second before or after the beginning and the end of a section, the media player can be a timesaver for working on large amount of data.

The transcriptions of source and target versions can be carried out separately and then merged automatically in one Partitur Editor file or they can be carried out simultaneously and then separated manually. In this sense, it is crucial to have ST and TT videos synchronised, so that spotting (in case of simultaneous transcriptions) and merging (in case of separated transcriptions) work smoothly. Furthermore, to make sure the process does not overload the software, it is important to keep a reasonable number of tiers. In the case of my framework, the number of tiers of the pilot was rather high, and the merging process did cause some slowing down. The work on the pilot showed the next step now being taken for the rest of the corpus: reduce the number of tiers, assigning more options to the tagging of each tier. For example, the category *discourse* is going to include also qǐng-like calques (请) and proverbiality.

3.2. Organisation of the AVT corpus using Corpus Manager

3.2.1. Description and uses for AVT studies

In Corpus Manager (CoMa) the transcriptions are organised together to become an organic corpus, as it allows to bundle transcriptions (with the respective multimedia files) and to add a series of metadata on different levels. It is optimised for bulky corpora. Here, a *communication* is a whole communicative event, including all its multimodal elements. The corpus is in fact organised in *communications* on the one hand and *speakers* on the other hand, and it is possible to associate the two with one another. Indeed, while the same speaker can be present in more than one communication, each communication usually has more than one speaker. The metadata entries can be attributed to the whole corpus, to the communication and/or to the speaker, and they can include all sorts of relevant information. The metadata list is flexible (can include anything from attached files, to predefined data, to free text/description), and the scholar can freely establish the categories.

As can be seen in Figure 3, on the left, circled in red, are the *communications* (at the moment there is only one, but there will be as many as the transcription files); on the right, circled in red, are the *speakers* (in this case, the ST and TT versions of the same character, Charlie Chan). Clicking on each communication and on each speaker allows to input and see the metadata. The sub-corpora can be created bundling communications and/or speakers. When a larger corpus is in place, the list will represent an excellent visualisation of the whole set of data.

https://doi.org/10.6092/issn.1974-4382/15557

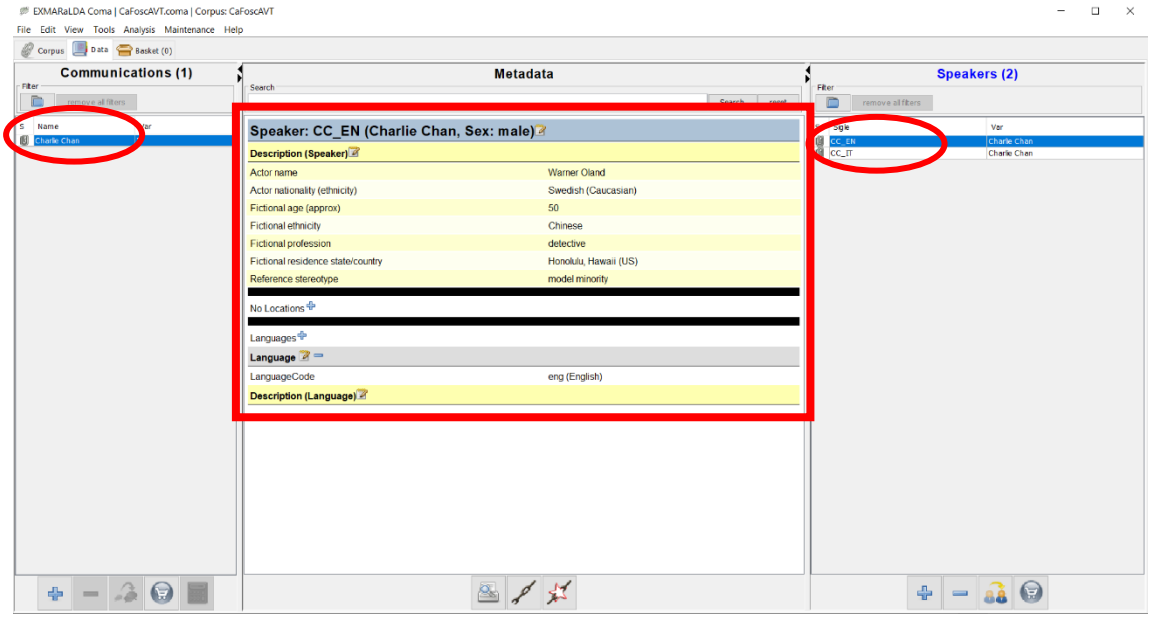


Figure 3. An example of speaker metadata in Corpus Manager (red square). Communications (left) and speakers (right) are circled in red.

One of the possible metadata levels (*character*) applied to my pilot study can also be seen in Figure 3 (in the red square). In particular, for Charlie Chan relevant metadata included information on both the character and the actor interpreting him. In this case, the Chinese character was interpreted by a Swedish actor wearing yellowface[11] on screen. The language category is "closed", as the compiler is asked to select the ISO 639-3 language code (e.g. eng for English, ita for Italian etc.) The categories in the yellow area (belonging to the *description* box) are instead open for the scholar to insert whatever information is deemed useful. Best practice, however, would be to have those rather standardised too. In fact, metadata are useful to categorise, but also to stratify the corpus. A solid and detailed metadata apparatus (with a consistent terminology) will allow responding to very specific research questions with relative ease. Using the metadata as filters, the scholar can also operate a selection of a number of communications/speakers and save them as sub-corpora. In the corpus I am building, a relevant example is sub-corpora of characters belonging to the same reference stereotype.[12]

With Corpus Manager it is possible to start a corpus from scratch – i.e. an empty file to be filled with transcriptions – or to automatically generate a corpus starting from transcriptions with a step-by-step guided procedure. One important action to carry out once the CoMa file is created is *segmentation*, an automatic

[11] Nowadays mostly abandoned, yellowface is, much like blackface, a discriminating practice consisting in onscreen make up aimed at reproducing East Asian features on the face of white actors, allowing them to play Asian characters. It was a common practice in the 1930s (Luise Rainer won an academy award donning yellowface in the 1937 film *The Good Earth*, McFadden 2014). Today it has mostly been replaced by what Ono and Pham (2009) call *implicit yellowface*, according to which Asian actors are asked to embody the image WASPs have of them, rather than their own experience.

[12] The reference stereotypes in this corpus include (but are not limited to) model minority, yellow peril, FOB (non-integrated immigrant), nerdy/eccentric, invisible (Renna 2021a).

https://doi.org/10.6092/issn.1974-4382/15557

operation that divides the transcriptions into utterances, words etc. This will create an additional version of the transcriptions with added information on tokenisation that will be used by the software itself. It is possible to generate a list of transcription errors found during software segmentation, so that the scholar can intervene in the Partitur Editor file to fix the issues. In fact, the error list is a special .xml file that can be open in Partitur Editor: there, clicking on each error will take to the place where it was found, simplifying troubleshooting (see Figure 4).

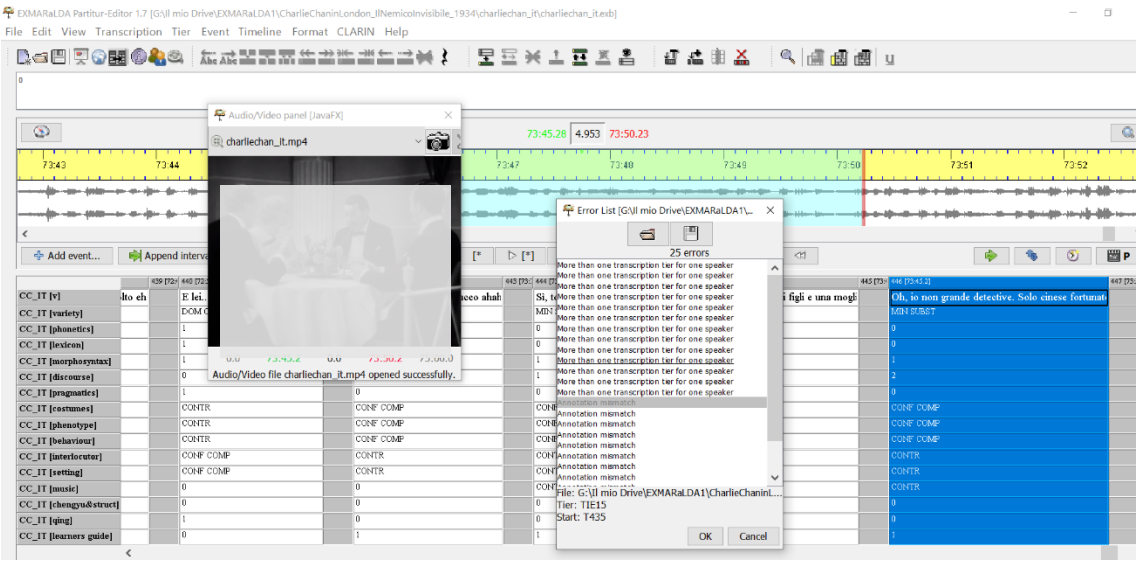


Figure 4. Error list generated in Corpus Manager and open in Partitur Editor for troubleshooting.

This function belongs to the *maintenance* menu, which contains some crucial CoMa functions. Since especially bigger corpora will be updated in the course of time, it is important to stress that Corpus Manager leaves the corpus "open" to updating and fixing at any moment, and this is all done through the *maintenance* menu, containing a series of options for both transcription/annotation and metadata. This allows automatising tedious aspects of corpus polishing to a great extent.

3.2.2 Outputs and results

The main output delivered by CoMa is the transformation of a series of multimodal transcriptions, separated and isolated from each other, in a comprehensive, searchable, updatable, and organised corpus (extension: .coma, file type: EXMARaLDA corpus). CoMa is where the corpus itself is visible as a whole.

Even though Exakt (the following step) is the tool specifically dedicated to querying, CoMa already contains some basic query tools, i.e. wordlist generation (general and by speaker), general corpus statistics and a reduced Exakt-like interface featuring part of the queries more extensively carried out in Exakt. In addition, CoMa allows part-of-speech tagging and querying, as it is linked to the software TreeTagger[13] via the *tools* menu. The TreeTagger interface will allow

[13] As can be read on its download webpage, TreeTagger is a free "tool for annotating text with

uploading a parameter file (the working parameters for several languages are found on the website) and an abbreviation file. The tagging is then carried out automatically, and the scholar can check and correct the results. Since ST and TT are usually different languages, the parameter files must be specific for each language; the best way to use TreeTagger in a bilingual or multilingual corpus is to split the different languages into sub-corpora and tag them separately. The part-of-speech tags thus obtained will be available for querying in Exakt.

It is worth noting that Corpus Manager performs some important organisational work "behind the scenes": one example is the aforementioned *segmentation*, which allows Exakt queries by automatically generating a "ghost" transcription. The scholar will not need to open this file: its role is to separate the parts of the transcription in a way that is recognisable by the software.

While transcription represents the data input moment, CoMa is the organising moment, in the same way one can write a series of books and then place and sort them in shelves based on different criteria, so as to ease the research of the titles (which is the following step).

3.3. A powerful tool for multimodal querying: using Exakt in AVT

3.3.1. Description and uses for AVT studies

Exakt is EXMARaLDA's concordance tool. Queries can be carried out on whatever was included in the Partitur Editor file, of course including both transcriptions and annotations, as well as on the CoMa metadata and the part-of-speech tagging, and can combine these various levels. Exakt can be used at any stage of corpus construction, so that the results can be queried and quantified *in itinere*.

The queries in Exakt are carried out through KWIC (keyword in context) concordance, which will probably be familiar to most scholars who have dealt with corpus-based research. KWIC queries in Exakt present the values of both transcriptions and annotations, but also display metadata and annotations in additional columns if needed. Exakt allows queries with regular expressions (Stubblebine 2007). It is possible to manually select/deselect and remove the results, as well as to annotate query results themselves. When a corpus is open in Exakt, the wordlist and the concordances will also be available (see the boxes to the left of Figure 4).

More queries can be carried out at the same time and opened in different tabs (similarly to browser tabs or MS Excel sheets). In the main query window, the search bar is at the top, where the scholar decides whether to start the search from transcriptions or annotations, and the buttons to the right give access to various options, such as filer application.

part-of-speech and lemma information […] developed by Helmut Schmid in the TC project at the Institute for Computational Linguistics of the University of Stuttgart", able to work with a series of languages, like "German, English, French, Italian, Danish, Swedish, Norwegian, Dutch, Spanish, Bulgarian, Russian, Portuguese, Galician, Greek, Chinese, Swahili, Slovak, Slovenian, Latin, Estonian, Polish, Persian, Romanian, Czech, Coptic and old French". https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (visited 19/07/2021). TreeTagger must be downloaded separately from its dedicated page.

https://doi.org/10.6092/issn.1974-4382/15557

To see more about a specific result, it is possible to click on it, and this will show the Partitur Editor full transcription and annotation, as well as the multimedia file(s).

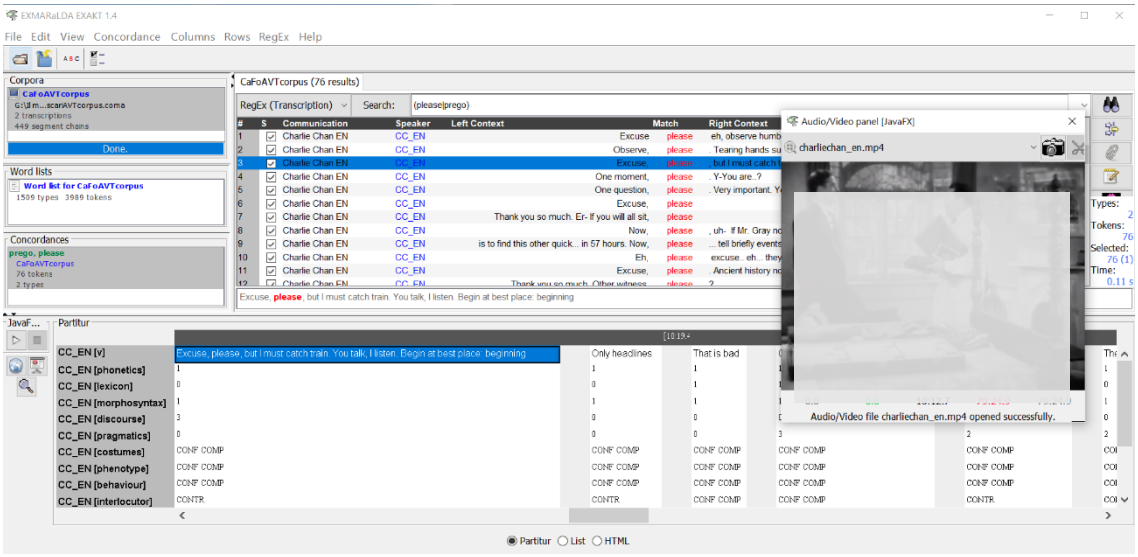


Figure 5. An example of transcription query with Exakt.

As can be seen in Figure 5, regular expressions were used to search for both ST and TT version of a word of interest simultaneously. As explained in Renna and Santulli (2022) the words *please* and *prego* were especially recurrent in the pilot study (in ST and TT respectively). Regular expressions that are useful for the scholar can be saved in a dedicated library within Exakt, to be easily retrievable for later queries.

If the transcriptions are merged, scrolling down on the Partitur Editor section will show both ST and TT versions of the selected line. Various combinations of transcription, annotation and metadata will be possible, so that the search can be adjusted and perfected according to the scholar's needs.

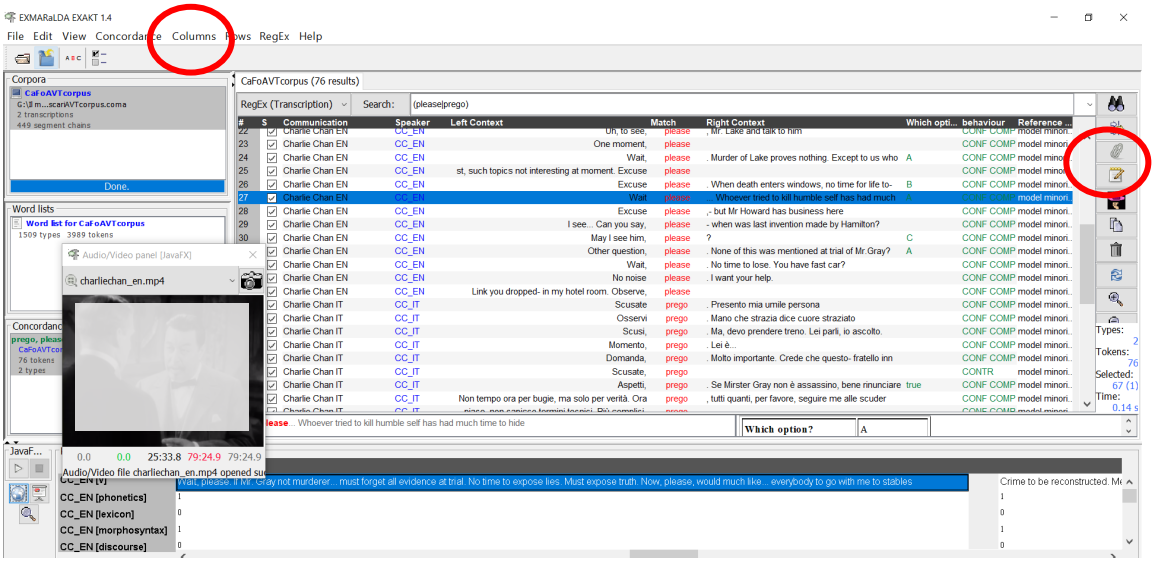


Figure 6. An example of crossed query in Exakt. The options to add more columns are circled in red.

https://doi.org/10.6092/issn.1974-4382/15557

Figure 6 shows how queried words (red), annotations and added analyses (both green) and metadata (black) are shown simultaneously, along with left and right context. Even more results can be added to the query. Here a sample query was carried out to show the interaction of different levels. The main query was based on the regular expression "(please|prego)", the same as Figure 5, but more layers were added. From the menu on the right (circled in red) I used both the *specify metadata* (paper clip) and the *add analysis* functions (notebook). The chosen metadata was *reference stereotype* (mentioned in the previous subsection) – since the pilot study consists in one character, the result is "model minority" in all cases. The *add analysis* function is flexible, as it is possible to add anything between binary options and free text. The sample analysis added in this case was a closed category list based on the question "Which option?", signalling which results corresponded to options deemed relevant during the query phase (if a certain category is often relevant across the corpus, the scholar may consider adding it in the transcription). Finally, using the *columns* menu (circled in red on the top left), I added one of the annotation categories, the multimodal relation between speech and behaviour, so as to visualise how many times the use of the queried words corresponded to a certain value in that relation, as well as to the reference stereotype and to the sample added analysis. All these options allow the scholar to personalise and adjust the query to obtain elaborate results.

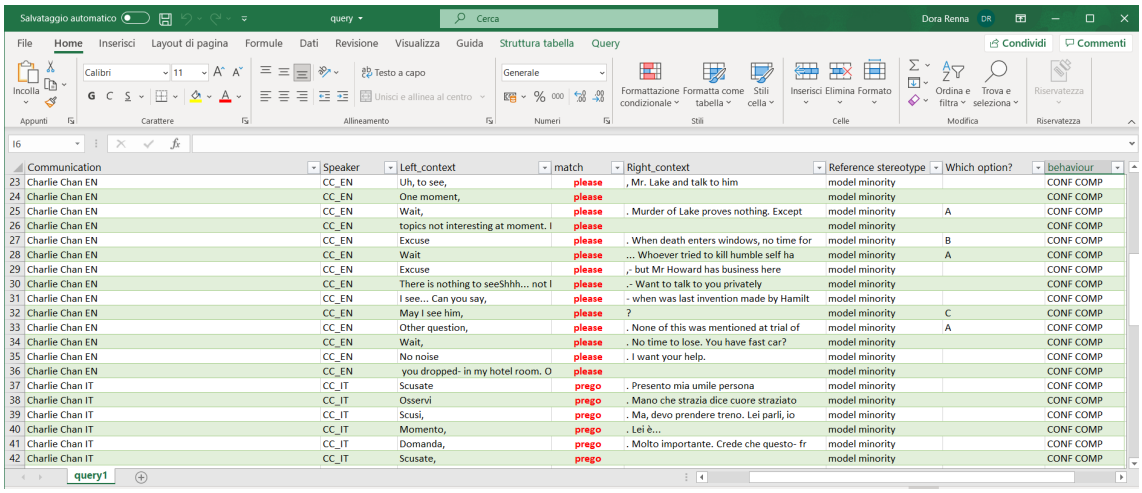
Exakt is arguably what makes EXMARaLDA stand out against competition the most. The queries are easily carried out, especially for those who are familiar with regular expressions. All levels of analysis are available for cross-analysis, which makes the whole process flexible, personalised, and multimodal. Having a merged transcription (which will correspond to alignment in this environment) allows the AVT scholar to see transcription, tagging and multimedia of ST and TT at the same time, at any point of their research.

3.3.2. Outputs and results

Exakt performs the final step of EXMARaLDA's work. It does not create tables and graphs, but allows very sophisticated queries that are visible and shareable, and can be then saved and retrieved, as well as exported into a statistics-specialised software like MS Excel (with which many scholars will probably be familiar) for graph generation. This can be done by saving the query as a .csv file, so that it can be opened in MS Office Excel to generate graphs and tables.

Figure 7 shows an example of Exakt query exported into MS Excel: all the categories are visible just like in Exakt (see Figure 6), and are ready for statistical elaboration and graph generation.

https://doi.org/10.6092/issn.1974-4382/15557



	Communication	Speaker	Left_context	match	Right_context	Reference stereotype	Which option?	behaviour
23	Charlie Chan EN	CC_EN	Uh, to see,	please	, Mr. Lake and talk to him	model minority		CONF COMP
24	Charlie Chan EN	CC_EN	One moment,	please		model minority		CONF COMP
25	Charlie Chan EN	CC_EN	Wait,	please	. Murder of Lake proves nothing. Except	model minority	A	CONF COMP
26	Charlie Chan EN	CC_EN	topics not interesting at moment. I	please		model minority		CONF COMP
27	Charlie Chan EN	CC_EN	Excuse	please	. When death enters windows, no time for	model minority	B	CONF COMP
28	Charlie Chan EN	CC_EN	Wait	please	... Whoever tried to kill humble self ha	model minority	A	CONF COMP
29	Charlie Chan EN	CC_EN	Excuse	please	,- but Mr Howard has business here	model minority		CONF COMP
30	Charlie Chan EN	CC_EN	There is nothing to seeShhh... not l	please	.- Want to talk to you privately	model minority		CONF COMP
31	Charlie Chan EN	CC_EN	I see... Can you say,	please	- when was last invention made by Hamilt	model minority		CONF COMP
32	Charlie Chan EN	CC_EN	May I see him,	please	?	model minority	C	CONF COMP
33	Charlie Chan EN	CC_EN	Other question,	please	. None of this was mentioned at trial of	model minority	A	CONF COMP
34	Charlie Chan EN	CC_EN	Wait,	please	. No time to lose. You have fast car?	model minority		CONF COMP
35	Charlie Chan EN	CC_EN	No noise	please	. I want your help.	model minority		CONF COMP
36	Charlie Chan EN	CC_EN	you dropped- in my hotel room. O	please		model minority		CONF COMP
37	Charlie Chan IT	CC_IT	Scusate	prego	. Presento mia umile persona	model minority		CONF COMP
38	Charlie Chan IT	CC_IT	Osservi	prego	. Mano che strazia dice cuore straziato	model minority		CONF COMP
39	Charlie Chan IT	CC_IT	Scusi,	prego	. Ma, devo prendere treno. Lei parli, io	model minority		CONF COMP
40	Charlie Chan IT	CC_IT	Momento,	prego	. Lei è...	model minority		CONF COMP
41	Charlie Chan IT	CC_IT	Domanda,	prego	. Molto importante. Crede che questo- fr	model minority		CONF COMP
42	Charlie Chan IT	CC_IT	Scusate,	prego		model minority		CONF COMP

Figure 7. Query from Figure 6 exported in MS Excel.

A thorough and detailed exploration of the pilot results and their sociolinguistic and translational implications is out of the methodological scope of this article, and can be found in Renna and Santulli (2022). However, it might be worth briefly relating on the specific results obtained using EXMARaLDA with this study. Thanks to the work with this software not only was it possible to answer the research questions (the graphs and tables were generated with MS Excel), but I was also able to "see" some further verbal and non-verbal phenomena that did not directly correspond with my research questions.

The first research question was: what fictional variety does the character speak in ST and TT? By querying the annotation tier [variety], I obtained a list of the varieties used by Charlie Chan, and I checked the relevance of the results. I saved the query and exported it to MS Excel, where I could then calculate recurrences: in both ST and TT, the highest-frequency variety was *minority substandard*, the most peripheral variety in the scheme (see Figure 1), with 73.5% and 73% respectively. The only difference worth mentioning was that, while in the TT there was a slight resort to dominant varieties (overall less 26.1% of the TT lines), ST Charlie Chan always verbally displayed his ethnic belonging. The following question was the one concerning variety realisation. In this case I queried three tiers simultaneously: [phonetics], [morphosyntax] and [lexicon] in both ST and TT, and then I restricted the query to the most recurrent variety (*minority substandard*). The query was subsequently exported in MS Excel for statistical elaboration. Here I noticed a crucial difference: in the ST morphosyntax and phonetic/suprasegmental traits are almost equally present (99.4% and 100% respectively), and lexicon appears marked in 35.5 % of the lines. The TT was significantly less marked, as phonetics dropped to as little as 2.4%, and lexicon decreased to 12.7%, but morphosyntax slightly increased, as it reached 100%. These data were crucial to answer the third question. While there are countless translational strategy taxonomies, I have opted for Ramos Pinto's (2017), designed for corpus-based study of language variation in AVT. Since the variety is mostly kept, but with less marked features, it is possible to infer that the most used strategy was *centralisation*, which "accounts for the cases in which the TT presents a lower frequency of nonstandard features (or the

https://doi.org/10.6092/issn.1974-4382/15557

choice for more prestigious features/variety in relation to the ST) and can thus be placed closer to the centre of prestige" (Ramos Pinto 2017: 9). The last question concerned intermodal relations. Here I used Exakt to query all the relation tiers ([behaviour], [costumes & makeup], [phenotype], [interlocutors], [setting] and [music]) simultaneously in ST and then in TT, I saved the queries and exported them into MS Excel. Data showed that in the ST confirmation relations are prevalent (70%), especially in the categories of phenotype, costumes and make up, and behaviour. Plot-required contradiction appears with interlocutors and setting, both representing the British élite. Music was nearly absent throughout the film and did not have statistical relevance. TT revealed a certain increase in contradiction (from 30% to 43%), but the confirmation/contradiction was overall kept.

As mentioned above, using EXMARaLDA helped me notice more phenomena than the ones I was looking for a priori. First, thanks to multimodal tagging I could "see" the single lines repeatedly with ease, and some very specific non-verbal features appeared recurring, so I decided to add them as transcription categories. The character has a distinctive and stereotypical proxemic behaviour, which mainly consisted in bows and joined hands (or both at the same time) and appeared in 65.9 % of his lines – a non-negligible frequency. This non-verbal display of exaggerated modesty is accompanied by verbal ones. These were noticed in Exakt, where I could query some words that seemed recurrent during transcription and annotation. In particular, among the verbal displays of modesty, it is worth citing the calque of the Mandarin word "qǐng" (请), which can be roughly translated with "please" and "prego" in English and Italian respectively. These two words were queried using regular expressions and turned out to appear in 17.7% of the ST and 21.2% of the TT lines. Another stereotypical discursive trait found in Charlie Chan was his proverbiality, especially realised with the use of similes, noticed during tagging and then queried in Exakt through a series of keywords such as "like". Proverbiality appeared in 19.2% of the ST and 17.7% of the TT lines in various forms: adaptation of real Chinese sayings, adaptation of chéngyǔ (成语, Chinese sayings composed by four characters), or made-up proverbs distinguishable thanks to their structure. While the percentages vary to some extent from ST to TT, in no case they disappear completely or increase/decrease to the point of doubling/halving– and this further confirms the inferred centralisation strategy.

4. Concluding remarks

While most scholars acknowledge the importance of multimodality and corpus-based approach in AVT, not all may decide to experiment with new software, especially if they come from qualitative approaches or if they already have a preferred software – albeit a non-multimodal one. This reluctance often depends on the overwhelming experience to be immersed in a completely new virtual environment. While approaching new tools can be discouraging and unnerving at first, many issues can be solved easily by seeking guidance and collaboration

https://doi.org/10.6092/issn.1974-4382/15557

with software developers and digital humanities experts. Mastering new techniques is an empowering effort, which pays off with more solid results.

In this sense, EXMARaLDA has so far proved to be a great tool, as it has been able to support the high level of complexity of an AVT scholar's research questions, which concern two languages simultaneously, as well as multimodality. Indeed, the software design allowed "the integration of different levels of linguistic description and different types of context data", while "permitting a two-way interaction between the data and its analysis" (Schmidt and Wörner 2009: 580). Regular software updating and fixing, and the availability of developers for support and training are also crucial, especially for those scholars who are not used to working with software.

There are some things that, during my training, emerged as pivotal for EXMARaLDA and, in general, for all software use in corpus-based studies. One is to deeply reflect on one's workflow: having in mind neat research questions is only the starting point – then it is necessary to become acquainted with the software, to fully understand how it can support the work throughout (rather than making it more complicated or slowing it down). This makes always worth trying a pilot study before deciding whether the initial choice of software and/or workflow is the most suitable for the objective. In this sense, the flexibility to adjust to emerging needs and even start again when necessary are researcher skills that become especially useful when working with new methodologies. A corollary to a clear workflow is an orderly management of the corpus material – which must be clearly organised, carefully stored and constantly backed up. Although this may seem obvious, it is quite hard to achieve especially as the amount of data increases. This is both true when working alone and in groups – each condition poses a series of different challenges linked to time consumption and shared practices.

The experimentation here described is still ongoing, and more will emerge as a full corpus comes to existence in EXMARaLDA – both in terms of methodology and results. I have now worked on similar projects with and without this software and, despite the initial hurdles of dealing with a completely new tool, there are significant advantages in shifting to a more digital approach. One is the actual multimodality of the whole work, from transcription to querying, which is linked to the features of Partitur Editor. Another is the fact that the features of an EXMARaLDA-generated corpus correspond with the requirements of my corpus in many respects, for example the *speaker* criterion for corpus sorting in Corpus Manager very well suited the focus on characters of my framework. Moreover, Exakt queries are easy to perform but at the same time can reach high levels of sophistication and complexity.

This work is an attempt to contribute to an increased synergy between translation studies (especially AVT) and digital humanities, and to point the way towards a deeper embracing of multimodal and corpus-based approach to AVT. The focus on character design was chosen because it is an aspect of the analysis that shows more than others the need to go beyond the textual dimension. In AVT, resorting to digital humanities can be the way forward towards embracing and delving into the polysemiotic nature (Chiaro 2008) of the audiovisual product.

https://doi.org/10.6092/issn.1974-4382/15557

ACKNOWLEDGEMENTS: I would like to thank Thomas Schmidt (University of Basel, Switzerland) for his patient guidance and training, and the Department of Linguistics and Comparative Cultural Studies at Ca' Foscari University of Venice for the support that made the training possible.

REFERENCES

Antonini, R. (2008) "The Perception of Dubbese: an Italian Study", in D. Chiaro, C. Heiss, and C. Bucaria (eds.) *In Between text and image: updating research in screen translation*, Amsterdam: John Benjamins Publishing Company, 135-148.

Bateman, J. A. (2014) *Text and image: a critical introduction to the visual/verbal divide*, London & New York: Routledge.

Chaume, F. (2004) "Film Studies and Translation Studies. Two disciplines at stake in audiovisual translation", *Meta* 49 (1): 12-24.

Chiaro, D. (2008) "Issues in Audiovisual Translation", in J. Munday (ed.) *The Routledge Companion to Translation Studies*, London & New York: Routledge, 141-165.

Díaz-Cintas, J. (2008) *The Didactics of Audiovisual Translation*, Amsterdam: John Benjamins Publishing.

ELAN - The Language Archive. https://archive.mpi.nl/tla/elan (visited 19/07/2021).

EXMARaLDA Official Website. https://exmaralda.org/ (visited 19/07/2021).

Feng, P. X. (2002) *Identities in Motion: Asian American Film and Video*, Durham & London: Duke University Press.

Fluck, W. (2003) "Film and Memory", in U. J. Hebel (ed.) *Sites of Memory in American Literatures and Cultures*, Heidelberg: Universitätsverlag C. Winter, 213-229.

Forde, E. (1934) *Charlie Chan in London [Il nemico invisibile]*.

Franklin, S., S. Wood, and G. Machatý (1937) *The Good Earth*.

Gambier, Y. (2006) "Multimodality and Subtitling", paper presented to the *Marie Curie Euroconferences MuTra: Audiovisual Translation Scenarios* (Copenhagen Copenhagen, 1-5 May 2006), https://www.euroconferences.info/proceedings/2006_Proceedings/2006_proceedings.html (visited 28/07/2021).

Gurskis, D. (2006) *The Short Screenplay: Your Short Film from Concept to Production*, Boston: Thomson Course Technology.

Hedeland, H., T. Lehmberg, T. Schmidt, and K. Wörner (2011) "Multilingual Corpora at the Hamburg Centre for Language Corpora". *Multilingual Resources and Multilingual Applications. Proceedings of GSCL Conference 2011 Hamburg*: 23-44.

Hill Collins, P. (2002) *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, London & New York: Routledge.

Huang, Y. (2010) *Charlie Chan: The Untold Story of the Honorable Detective and His Rendezvous with American History*, New York: W. W. Norton & Company.

Iedema, R. (2003) "Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice", *Visual Communication* 2 (1): 29-57.

Laviosa, S. (2002) *Corpus-based Translation Studies. Theories, Findings, Applications*, New York: Rodopi.

Lê Espiritu, Y. (1992) *Asian American Panethnicity: Bridging Institutions and Identities*, Philadelphia: Temple University Press.

Lemke, J. (1998) "Multiply meaning: Visual and verbal semiotics in scientific text", in J. R. Martin, and R. Veel (eds.) *Reading science: critical and functional perspectives on discourses of science*, London & New York: Routledge, 87-113.

Lowe, L. (1991) "Heterogeneity, Hybridity, Multiplicity: Marking Asian American Differences", *Diaspora* 1(1): 24–44.

Ma, S.-m. (2014) *Alienglish. Eastern Diasporas in Anglo-American Tongues*, Amherst: Cambria Press.

McFadden, R. D. (2014) "Luise Rainer Dies at 104; Won Best Actress Oscars for Two Years Running", *The New York Times*. https://www.nytimes.com/2014/12/31/arts/luise-rainer-award-winning-actress-dies-at-104.html (visited 25/07/2021).

McLuhan, M. (1964) *Understanding Media: The Extensions of Man*, New York: McGraw Hill.

Naremore, J. (1988) *Acting in the Cinema*, Los Angeles: University of California Press.

Ono, K. A., and V. N. Pham (2009) *Asian Americans and the Media*, Cambridge: Polity Press.

Pastra, K. (2008) "COSMOROE: a Cross-Media Relations Framework for Modelling Multimedia Dialectics", *Multimedia Systems* 14: 299-323.

Pérez-González, L. (2014) *Audiovisual translation: theories, methods and issues*, London & New York: Routledge.

Praat: Doing Phonetics by Computer. https://www.fon.hum.uva.nl/praat/ (visited 19/07/2021).

Ramírez Berg, C. (2002) *Latino images in film: stereotypes, subversion, and resistance*, Austin: University of Texas Press.

Ramos Pinto, S. (2009) "How important is the way you say it? A Discussion on the Translation of Linguistic Varieties in Different Media". *Target: International Journal of Translation Studies* 21 (2): 289-307.

------ (2018) "Film, dialects and subtitles: an analytical framework for the study of non-standard varieties in subtitling", *The Translator* 24 (1): 17-34.

------ and A. Mubaraki (2020) "Multimodal Corpus Analysis of Subtitling: The Case of Non-standard Varieties", *Target: International Journal of Translation Studies* 32 (3): 389-419.

Renna, D. (2021) *Language Variation And Multimodality In Audiovisual Translation – A New Framework Of Analysis*, Hannover: ibidem-Verlag.

------ (2021a) "Shaping Stereotypes in a Space of Absence. A Linguistic Analysis of Wayne Wang's Chan Is Missing". *Iperstoria* 17: 115-136.

------ and F. Santulli (2021) "*Me, lowly Chinaman*: analisi linguistica, multimodale e pragmatica dell'eloquio di Charlie Chan e del suo doppiaggio italiano", in

G. E. Garzone, E. Liverani (eds.) *Tradurre l'oralità. Aspetti pragmatici e culturali, mediAzioni* 31: A289-A316.

Schmidt, T., and K. Wörner (2009) "EXMARaLDA - Creating, Analysing and Sharing Spoken Language Corpora for Pragmatic Research", *Pragmatics* 19(4): 565-582.

Stubblebine, T. (2007) *Regular Expression Pocket Reference: Regular Expressions for Perl, Ruby, PHP, Python, C, Java and .NET*, Sebastopol: O'Reilly.

TreeTagger - A Part-of-Speech Tagger for Many Languages. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (visited 19/07/2021).

van Doorslaer, L., P. Flynn, and J. Leersen (eds) (2016) *Interconnecting Translation Studies and Imagology*, Amsterdam: John Benjamins Publishing.

https://doi.org/10.6092/issn.1974-4382/15557