# Issues in the Multilingual Information Processing of Spoken Political and Journalistic Texts

# Issues in the Multilingual Information Processing of Spoken Political and Journalistic Texts

By

Christina K. Alexandris

Cambridge
Scholars
Publishing

Issues in the Multilingual Information Processing of Spoken Political and Journalistic Texts

By Christina K. Alexandris

This book first published 2020

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library



ISBN (10): 1-5275-4510-5
ISBN (13): 978-1-5275-4510-6

# TABLE OF CONTENTS

# CHAPTER ONE

# INTRODUCTION: THE DOMAIN OF SPOKEN JOURNALISTIC TEXTS AND THEIR CHALLENGES

## 1.1 Introduction: Information Understanding, Information Transfer and Cognitive Bias in Spoken Journalistic Texts

Spoken political and journalistic texts are a commonly occurring spoken text type in the Media, available to recipients of almost any kind, ranging from a local community to the international public. Beyond television networks and broadcast news, spoken political and journalistic texts are accessible and retrievable with multimedia applications, viewed on a personal computer (PC), a mobile phone or on other devices.

In addition to their ubiquity, spoken political and journalistic texts constitute a significant source of empirical data both for human behaviour and for linguistic phenomena, especially for spoken language.

However, with some exceptions, spoken political and journalistic texts are usually underrepresented both in linguistic data for translational and analysis purposes and in Natural Language Processing (NLP) applications.

As text types, spoken political and journalistic texts pose challenges for their evaluation, processing and translation due to a set of typical, distinctive characteristics. Spoken political and journalistic texts are usually rich in socio-linguistic and socio-cultural elements. Additionally, the content of most spoken political and journalistic texts is often not domain-specific but may encompass several domains. Furthermore, with spoken political and journalistic texts there is always the possibility of different types of target audiences – including the international community. These features of spoken political and journalistic texts may often create complications with respect to the correct understanding of their information content or the correct transfer/ translation of their information content to a different audience.

An additional feature of spoken political and journalistic texts is the existence of complex and implied information often containing indications of the Speaker's attitude and intentions. This information is not always perceived or correctly understood by the recipients, particularly if an international public is concerned. This type of information is also not easily detectible by most Opinion Mining applications. Furthermore, information content and its perception by the recipients is often related to Cognitive Bias.

Additionally, the relation of the spoken information content to prosodic and paralinguistic features is a distinctive element of spoken political and journalistic texts which needs to be integrated in any evaluation, translation or other processing tasks performed, either by a human or by a Natural Language Processing (NLP) application.

Here, we focus on the basic, typical, characteristic and most commonly occurring issues and problems in the evaluation, processing and transfer of information content of spoken political and journalistic texts and present the main task types, general characteristics and basic factors to be taken into consideration for strategies targeting the resolution of the problems and the management of the issues concerned.

Finally, the signalization and visualization of information that is mostly not uttered constitutes a typical issue that can pose challenges to applications processing spoken journalistic texts. We present strategies and solutions such as the visualization of the degree of neutrality of a written or spoken journalistic text with the use of percentages; the visualization of Speaker behaviour, Speaker intentions and Cognitive Bias with the use of graphs; and the visualization of information content of paralinguistic features with the use of messages.

All proposed strategies implemented and presented are based on the Gricean Cooperative Principle (Grice, 1975) for the speech acts involved. The Gricean Cooperative Principle (Grice, 1975) serves as a general theoretical framework at the pragmatic level: parameters may vary according to the natural languages and cultural features involved, especially if non-European languages are concerned. We focus on English as an international language and related or comparable issues and linguistic phenomena in other languages.

Proposed approaches and strategies are presented taking into account the "grey zones" – undefined boundaries between individual-specific, language-specific and socio-cultural-specific parameters of spoken political and journalistic texts.

The proposed approaches and strategies intend to address the following issues concerning the processing of these texts:

- How to indicate and formulate the factor concerning the possibility of different types of domains and/or target audiences of the spoken journalistic text and how to integrate this factor for analysis and processing purposes.
- How to indicate and formulate the factor of speaker attitude and speaker intentions and how to integrate this factor for analysis and processing purposes.
- How to indicate and formulate the factor of prosodic and paralinguistic information and how to integrate this factor for analysis and processing purposes.

## 1.2 Challenges of Processing Spoken Journalistic Texts

The challenges of spoken political and journalistic texts concern three main task types performed either by humans or by NLP applications, namely (1) evaluation, (2) processing and (3) transfer of information.

These challenges are related to the following general characteristics of spoken political and journalistic texts:

(i) multiple domains;
(ii) multiple (national and international) audiences;
(iii) multiple language styles (for example, standard language and mixed terminology, possibly along with idiomatic, dated or scholarly language);
(iv) implied information and connotative features; and
(v) prosodic and paralinguistic features.

- **Multiple Domains**

Factors that plays an important role, particularly when it comes to translation, machine translation and Natural Language Processing (NLP) applications are the multiple domains and the (international) public's varied levels of knowledge in respect to terminology. Spoken journalistic texts involve a broad spectrum of terms and expressions, ranging from the domains of Politics and Diplomacy (including military terminology) to domains involving specialized terminology from other professional, technical and scientific fields such as Information Technology (IT), Engineering, Chemistry and Medicine (Tassis and Iliakis, 2015).

- **The Domain of International Affairs**

For example, in the domain of International Affairs, encompassing international politics, geopolitical and military affairs, texts often contain terminology from more than one specialized domain, resulting in ambiguities and other complications in translation, Machine Translation tools and other Natural Language Processing applications. We note that accuracy is of crucial importance in Machine Translation applications – a still-evolving technology – where the impact of mistranslations does not only effect critical, "life-or-death" situations in medical and technical texts, but also has impacts in the domains of Politics and Diplomacy.

For the domains of geopolitical and military affairs, the importance of information processing, including multilingual Natural Language Processing (NLP) – especially in the domain of military applications – has been described both by researchers and military personnel, including cases where attempts "to develop policy and create new programs were often lost in translation" (West, 2009). The processing of documents in different languages may be particularly problematic in applications concerning the analysis of textual data for intelligence purposes, especially if very few or no translators are available (Noubours and Hecking, 2012). Typical Natural Language Processing applications in the domains of International Politics and Geopolitical and Military Affairs include Information Retrieval (IR), Document Classification, Information Extraction (IE), Text Mining, Opinion Mining and Machine Translation (MT) (Noubours and Hecking, 2012). However, the nature and related complexity of spoken political and journalistic texts may sometimes not allow a clear outline and a distinction between features of their content and structure. This property may create complications in the evaluation, analysis and processing, including for the above-mentioned typical Natural Language Processing applications.

- **Complex and Implied Information**

An additional factor for the analysis and processing of spoken political and journalistic texts is the existence of complex and implied information often containing indications of the Speaker's attitude and intentions, related to language-specific parameters and socio-linguistic and socio-cultural elements. Complex and implied information also concerns possible combinations of features both from the native language and from the natural language spoken, when speakers from the international community are involved. Complex and implied information is not easily managed by

most Machine Translation tools or by most Natural Language Processing (NLP) applications.

- **Prosodic and Paralinguistic Features**

Furthermore, in spoken political and journalistic texts, a connection of the information content of spoken utterances to prosodic and paralinguistic features is observed. The factor of the Prosodic and Paralinguistic Level may also be connected to language-specific parameters and socio-linguistic and socio-cultural elements.

- **Transcription and Proposed Approaches**

Transcription tools used for spoken texts and speech applications facilitate the evaluation, analysis and processing of spoken political and journalistic texts, especially in interactive applications allowing the user to make decisions concerning the spoken input.

Proposed strategies and interactive annotation approaches in combination with existing automatic processes, tools and resources may function as "stepping stones" and "safety nets" (Lewis, 2009; Lewis, 2010) for the correct and complete evaluation, processing and/or transfer of the information content of spoken political and journalistic texts while "respecting" the nature and complexity of the processed spoken input.

The proposed strategies and interactive annotation approaches are related to their integration in implemented or designed applications such as the following, among others:

- Interactive processing of terminology in journalistic texts adapted to the needs of the international public (with use of existing resources).
- Automatic evaluation of the degree of "non-neutral" elements (degree of neutrality-objectivity) in journalistic texts.
- Graphic representations of content and quality of spoken interactions (interviews, discussions).
- Visibility and "processability" of prosodic and paralinguistic information for multilingual applications.
- Connection and accessibility to the international public of online spoken journalistic texts in respect to related content of ancient or historical texts.

- **Targets of Present Analysis and Strategies**

The presented and described features of spoken political and journalistic texts and the proposed strategies and interactive annotation approaches concern the following targets:

Target 1: Visibility of information implied and/or information not uttered and visibility of Speaker intentions and attitude.

Target 2: A general framework for both the correct and efficient analysis and/or processing (A) by human evaluators, analysts or translators and (B) by Natural Language Processing (NLP) applications, including Machine Translation and Data Mining (Opinion Mining/Sentiment Analysis).

Target 3: Accessibility of information content and correct information transfer for non-native Speakers and/or an international audience.

# CHAPTER TWO

# ISSUES IN PROCESSING SPOKEN POLITICAL AND JOURNALISTIC TEXTS

## 2.1 Visibility and "Processability" of all types of Information Content for Evaluation, Translation, Data Mining and Opinion Mining

The present analysis and proposed strategies may be described as a general framework with a "three level" processing approach described in the present chapter. The "three level" processing approach presented here targets the visibility of all features and respective information content of the spoken political and journalistic texts. This includes visibility of implied information and connotative features not evident at word level; information at the Prosodic and Paralinguistic Level; and information not uttered. Visibility of all information content ensures the "processability" of all information content.

Processability includes Data Mining and Opinion Mining applications that mostly rely on word groups and word sequences, including recent approaches with the use of neural networks (Arockiaraj, 2013; Shah et al., 2018). In particular, Opinion Mining applications make use of sentiment lexica (Liu, 2012), although recently neural networks have been employed in unannotated texts to improve performance (especially in low-resource languages) and to overcome errors (Hedderich and Klakow, 2018).

It should be noted that most Opinion Mining and Sentiment Analysis applications concern domains such as customer reviews and/or social media posts – even videos (Poria et al., 2017) – as illustrated in the following characteristic examples (Examples 2.1.1, 2.1.2 and 2.1.3).

Example 2.1.1

Product reviews for Sentiment Analysis and Opinion Mining (retail) (Liu, 2012):

(i)

"I bought an iPhone two days ago.
It looks very nice.
I made many calls in the past two days.
They were great."

(ii)

Posted by: John Smith Date: September 10, 2011

"(1) I bought a Canon G12 camera six months ago.
(2) I simply love it.
(3) The picture quality is amazing.
(4) The battery life is also long.
(5) However, my wife thinks it is too heavy for her."

Example 2.1.2

Movie reviews with User-generated videos (Poria et al., 2017):

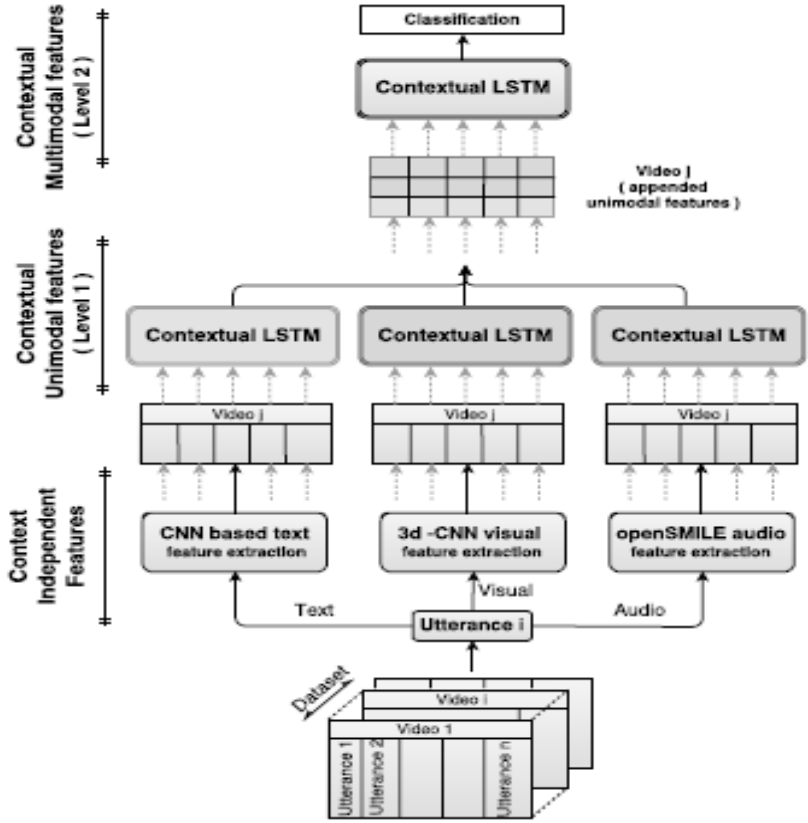
"What would have been a better name for the movie".
"And I really enjoyed it" –
"The countryside which they showed while going through Ireland was astoundingly beautiful"

Example 2.1.3

Sentiment Analysis from videos (text, audio and video):

Hierarchical architecture for extracting context-dependent multimodal utterance features (with Long Short-Term Memory (LSTM) recurrent neural networks) (Poria et al., 2017).

(Poria et al., 2017)




These domains are related to a variety of issues that analysts and developers must resolve. However, spoken political and journalistic texts pose additional challenges, as described in the present chapter. The nature and complexity of the content of spoken political and journalistic texts

renders the use of specially annotated corpora a necessity, at least as initial training and test sets for any application.

The evaluation, (machine) translation and other forms of processing of spoken journalistic texts concern the complexity of the texts in regard to the existence of multiple sub-domains involved (i) and in respect to linguistic processing (ii) (Complexity –A). The degree of complexity (the Complexity Issue – A) increases where multilingual texts from diverse languages and language families and an international public are concerned. Additional issues are processing speed (Speed – B) and availability of language resources (the Resources Issue – C). In particular, the efficient use of resources (the Resources Issue) and processing speed (the Speed issue) play an important role, especially if decision-making is involved.

## 2.2 Issues in Processing Spoken Journalistic Texts: Complexity, Speed and Resources

### 2.2.1 The Issues of Complexity and Speed

The complexity of linguistic processing in spoken political and journalistic texts (Complexity – A) includes the basic problems of (1) ambiguity resolution, (2) the ability to detect and process complex information and (3) access to related and/or implied information and connotative features. Spoken journalistic texts contain problems related to the semantics of words, such as ambiguities and polysemy; complex information from various domains (including International Affairs and Diplomacy); and implied information and connotative features contained in texts reflecting tendencies, opinions and positions taken in respect to events.

Additional issues concern the detection and processing of information that is not easily detected, often resulting in complications in applications such as Machine Translation and Data Mining (including Opinion Mining and Sentiment Analysis).

Ambiguity resolution and complex information processing are related to the multiple sub-domains that exist in spoken journalistic texts. Thus, the determination and application of domain-specific resources (terminology databases, lexica and corpora) with the ability to connect to multiple domains constitutes a basis for the efficient machine translation and processing of spoken political and journalistic texts.

Furthermore, the need for successful analysis and processing, a basic requirement in all types of applications, is coupled with the factor of urgency and the need to make an immediate decision, a common case in many types of (written and) spoken political and journalistic texts. Thus,

processing speed (Speed – B) constitutes another basic issue for processing texts involving typical domains of journalistic texts such as International Politics and Geopolitical and Military Affairs.

### 2.2.2 Using Available Resources

The access and effective use of resources (Resources – C) is an additional issue in the translation, Machine Translation and other forms of processing spoken journalistic texts. Bilingual and multilingual resources such as Parallel Corpora play a key role in the processing concerned. Typical examples of Parallel Corpora are the EuroParl corpus (European Parliament Proceedings Parallel Corpus, http://statmt.org/europarl/) and the United Nations Parallel Corpus (https://conferences.unite.un.org/uncorpus). An example of an open, freely available online parallel corpus is OPUS (http://opus.nlpl.eu/), providing aligned texts in multiple languages and connected to corpora including less widely spoken languages such as SETIMES2, a parallel corpus of news articles in the Balkan languages (http://opus.nlpl.eu/SETIMES2.php). The UM-Corpus (UM-Corpus: A Large English-Chinese Parallel Corpus http://nlp2ct.cis.umac.mo/um-corpus/) is an example of a domain-specific parallel corpus for the domains "Education", "Science", "Laws", "Thesis", "Microblog", "News", "Spoken" and "Subtitles".

In the strategies presented in the following chapters, readily available resources save time and reduce the cost of developing new corpora and multiple databases. The use of resources (Resources – C) in processing spoken journalistic texts concerns the combination of domain-specific/sublanguage-specific terminology and existing or specially constructed resources such as translations, texts, ontologies and other types of databases.

Specifically, in addition to corpora, the strategies may also include lexical databases such as WordNet (https://wordnet.princeton.edu/), multilingual databases such as the Universal Networking Language (UNL) (http://www.undl.org/), constructed monolingual or multilingual ontologies (e.g. OWL, Loaiza et al., 2014) or even monolingual lexica or corpora of languages from diverse language families such as Chinese or Arabic. Less spoken (local) languages may be processed in combination with "assistive" translations in a third language, with an option for further improvement and "fine-tuning" of translations.

These resources contribute to the processing of monolingual or multilingual journalistic texts for an international or national audience.

## 2.3 Overview of Existing Tools and Resources for Machine Translation

### 2.3.1 Recent Advances in Machine Translation and the Issue of Accuracy

#### 2.3.1.1 Introduction

Targeting the increasing needs of a multilingual international public, along with the needs of industry, government, and a variety of sectors and user groups, many Machine Translation (MT) systems continue to use a combination of various machine translation approaches.

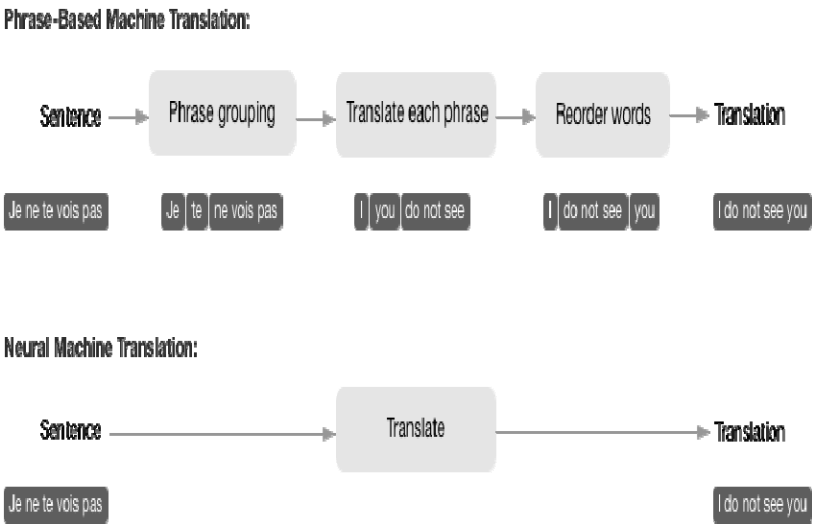
Statistical machine translation (SMT) (Koehn, 2010) has become increasingly popular compared to the original approaches of rule-based machine translation (RbMT) and even in comparison with example-based machine translation (Turcato and Popowich, 2003).

Statistical machine translation relies on pre-translated parallel corpora (Resources) on which the derived statistical probabilities are based. Although it dates back to the 1980s, rule-based machine translation, relying on hand-written grammar rules and lexica with vocabulary lists, is still used in some types of applications.

Despite recent developments, whether for the use of individual users or for industry, Machine Translation continues to vary in terms of levels of output accuracy (Burchardt et al., 2017), although in many cases better results are observed for neural machine translation (NMT) in comparison to phrase-based translation (PBNT), in terms of statistical machine translation approaches (Example 2.3.1).

Example 2.3.1

Outline of the phrase-based translation (PBNT) and the neural machine translation (NMT) processes (Zhou, Kurenkov and See, 2018)[1]




A characteristic illustration of these differences was made available to a broader public through Google press releases in 2016 concerning the performance of Google Translate.[2] However, even in the case of neural machine translation, researchers have used statistical machine translation to further improve translation results, "alleviating limitations" detected in neural machine translation processes (Wang et al., 2017).

Despite the above-presented advances, in terms of output accuracy, the performance of human translators continues to exceed all types of machine

[1] https://www.skynettoday.com/editorials/state_of_nmt
Skynet Today, July 25, 2018
Putting AI News in Perspective. Has AI surpassed humans at translation? Not even close! Neural network translation systems still have many significant issues which make them far from superior to human translators, by Sharon Zhou, Andrey Kurenkov and Abigail See

[2] https://www.theverge.com/2016/9/27/13078138/google-translate-ai-machine-learning-gnmt (by Nick Statt, September 27th, 2016)
https://www.techemergence.com/machine-translation-14-current-applications-and-services (by Radhika Madhavan, July 31st 2018)

translation, including neural machine translation (NMT) (Dietterich, 2017).

The recently launched Google Neural Machine Translation system (GNMT) (Wu et al., 2016) uses neural machine translation technology based on the use of a large neural network. Unlike other common Machine Translation procedures where the input is segmented into words and phrases as single units for translation (Jurafsky and Martin, 2008), in neural machine translation technology, the single unit for translation is considered to be an input sentence. The large neural network used predicts how likely a given sentence as a word sequence is to occur. The neural network encodes the source language words as a list of vectors. The meaning of all words read by the system so far is represented with each vector. The target language sentence is generated one word at a time, ending with the reading of the entire sentence by the system. In neural machine translation, entire sentences are modelled in a single integrated model (Cho et al., 2014).

In 2016, Facebook launched its own AI-powered translation system, which, according to press releases, marks the complete transition from phrase-based models into neural machine translation (NMT).[3] An additional recent development of neural machine translation is deep neural machine translation, which uses multiple neural network layers.

Neural networks are used in the Microsoft Translate technology for the Skype voice translation feature, Skype Translator, processing languages such as English and Spanish (https://www.skype.com/en/features/skype-translator/). The Skype voice translation software was initially based on a combination of speech recognition and statistical machine translation.

In speech-to-speech translation, the source audio is converted into raw text in the natural language spoken in the Automated Speech Recognition (ASR) module. The text normalization procedure processes the text for conversational speech to facilitate translation with the text translation engine. The text translation engine is built on specially developed translation models for spoken conversations in real-life situations.

There have been various demonstrations of Skype's successful speech-to-speech translation in real-time, including independent project videos. However, reservations have been expressed in respect to a successful implementation in online classrooms and business conversations between

[3] https://techcrunch.com/2016/05/23/facebook-translation/ Facebook ditches Bing, 800M users now see its own AI text translations, Josh Constine, (https://www.wired.com/2017/05/facebook-open-sources-neural-networks-speed-translations/, 05.09.17 Facebook's New AI Could Lead to Translations That Actually Make Sense, Cade Metz

organizations and in implementations that cross entity boundaries. These concerns involve both practical and security issues.

### 2.3.2 Domain-Specific Texts and the Machine Translation Market

Accuracy levels vary between language pairs and also in respect of the text types translated. It has been observed that generic translators with machine-learning models are trained on generic data and usually demonstrate a lower accuracy and relevancy level than domain-specific translators based on highly-specific training data.

In general, adaptation to the domain-specific needs of the user is typical of most machine translation services available in the market. These machine translation services, a significant percentage of which constitute Computer-Aided Translation (CAT) tools (Trados being a characteristic example of the oldest companies in the market providing such tools, https://www.sdltrados.com), are based on client-specific translation memory tools and customized machine translation engines. Applications such as GlobalLink are geared towards enterprise level translation and localization tools (http://translations.com/globallink/products/globallink_modules.html) including features and specifications for managing "highly branded, technical, or industry-specific terminology".

On the other hand, freelancers providing translation services are included in the main target group of machine translation software such as Lionbridge Translation Workspace (https://geoworkz.com/). Other machine translation services such as Andovar (https://www.andovar.com) place special emphasis on translation and localization, for example, languages of Southeast Asia.

Nevertheless, there are machine translation services available in the market for specialized machine translation applications. For example, in the domain of Finance, the Lingua Custodia Machine Translation system (http://www.linguacustodia.finance/en/homepage/) processes financial documents and is customizable to specific sub-types of financial documents, such as fund annual reports and portfolio management commentaries, with the system's ability to "learn" from previously translated texts. In the domain of Healthcare, Canopy Speak (https://withcanopy.com/speak/) is an example of a medical translator application with phrase-based translation, having a phrase library constituting a corpus of pre-translated medical phrases, categorized according to type and frequency of medical condition and medical procedure.

In the case of multi-domain translation services, the pre- and post-editing processes are still performed by human translators. Multi-domain translation services are usually offered across a defined number of specific domains, for example, the SYSTRAN Machine Translation system offers a variety of translation services for five domains/industries (http://www.systransoft.com/lp/machine-translation/). Multi-domain translation services may also involve a customization of their Machine Translation platform for specific domains, as in the case of the KantanMT Machine Translation system (https://www.kantanmt.com/).

### 2.3.3 Machine Translation of Texts, Audio Files, Images and Social Media Feeds

Despite their non-domain-specific application and their training on generic and typically crowd-sourced data, a high translation accuracy has been reported for commercial online cloud-based machine learning applications (apps) performing instant translation of textual, audio and image files for individual users. These machine learning applications offer Text-to-Text, Text-to-Speech, Speech-to-Text, Speech-to-Speech and Image-to-Text translation services.

An example of real-time Speech-to-Speech Translation is iTranslate Voice 2 (iTranslate), a speech-to-speech translation application (app) for travellers and tourists (http://itranslatevoice.com/). The iTranslate Voice 2 speech-to-speech translation app supports 42 languages. Its voice recognition and machine translation software interacts with a "Phrasebook" feature, allowing quick access to frequently used phrases. Other examples of similar speech-to-speech translation apps for travellers and tourists are SayHi, Speak and Translate, Papago and TripLingo, including wearable Speech-to-Speech Translation devices. A typical example is the Pilot (Waverly Labs) wearable Speech-to-Speech Translation device, consisting of a pair of wireless earbuds for facilitating a two-way conversation between speakers of two different languages (https://www.waverlylabs.com/company/). The application currently supports four Romance languages and English. Another example is the "ili" voice translator (https://iamili.com/us/), a small, handheld device voice translator for the translation of simple, common phrases for travellers, without the use of the internet, since it operates offline. It is based on a built-in phrase-based translation engine. However, the translation engine can be updated by plugging the device into a USB port and connecting to the company's website. The device processes the languages English, Spanish, Mandarin Chinese, and Japanese and only performs a one-way translation since,

according to the company's website, "travellers and tourists typically ask simple phrases that require yes or no answers".

An example of real-time Image-to-Text Translation is Textgrabber (ABBYY), translating the text from digitized images from any surface in "almost" real-time with the use of optical character recognition (OCR) technology (http://www.textgrabber.pro/en/).

The machine translation of social media feeds, documents, audio, and images in multilingual content and in real time has been described in presentations of services concerning the US government, focusing on defence and intelligence (intel) use cases, as in the case of the company SDL Government (http://www.sdlgov.com/). The translated content is subjected to further evaluation, for example, whether a posted social media feed indicates individuals who might potentially pose a national threat, as described by SDL Government's CEO Danny Rajan.[4]

### 2.3.4 Issues in Managing Machine Translation

Accuracy level is the key issue in machine translation applications, constituting the criterion for evaluating their success and efficiency. Inaccurate information and mistranslations continue to exist in machine translation output, not only because machine translation systems are a still-evolving technology, but also due to the complexity of the phenomenon of language itself.

As previously mentioned, an additional issue in machine translation applications is the consequences and impact of mistranslations, not only in the case of critical, "life-or-death" situations, such as in the case of technical texts and medical texts, but also in the domains of Politics and Business.

A third issue is the issue of security and the safe processing of confidential data. Security may concern personal data of users or confidential data of companies, organizations, government or other entities.

Table 2.3.1 summarizes the issues in managing Machine Translation in respect to common categories of available Machine Translation tools for spoken political and journalistic texts.

[4] interview Jun 17, 2017, https://www10.giscafe.com/video/SDL-Government-S.-Danny-Rajan-CEO/411334/media.html

**Table 2.3.1 Issues in managing Machine Translation and common categories of available Machine Translation tools for spoken political and journalistic texts**

| Common categories of available Machine Translation tools for spoken political and journalistic texts | Issues in Managing Machine Translation |
|---|---|
| specialized / domain-specific machine translation applications<br><br>multi-domain translation services<br><br>real-time Speech-to-Speech Translation<br><br>real-time Image-to-Text Translation | accuracy level<br><br>consequences and impact of mistranslations<br><br>confidential data |

## 2.4 Overview of Existing Tools and Resources for Transcription

The choice of transcription tools is usually dictated by the targeted form and type of the generated transcription; however, practical issues may play an additional role.

Easy access to transcription services for a large user group is provided by free online transcription tools such as YouTube and Google Docs. A characteristic example of an online transcription service for UK English audio recordings is "Standish Typing Services Transcriptions" (https://standishtyping.com/). Transcription tools can be customized for specific domains; for example, texts in the domain of the Law, with the transcription software of Dragon Speech Recognition Software (https://www.nuance.com/dragon.html) being one characteristic example.

The range of available transcription tools covers applications providing the basic requirements of text transcription and also more sophisticated software presenting multiple additional features such as linguistic information, prosody and gestures.

A typical example of the latter type of transcription tool is the freely available Anvil Video Annotation Research Tool (http://www.anvil-software.org/), which presents transcribed texts aligned with the respective video and speech signal as data imported from phonetic tools such as Praat. Praat is widely used software for the phonetic analysis of speech, mostly used for research and development in the field of articulatory and speech synthesis (http://fonsg3.hum.uva.nl/praat/).

Other characteristic examples of available online transcription tools with additional specifications are the recently released Google Live Transcribe speech engines (https://www.android.com/accessibility/live-transcribe/), the EXMARaLDA (http://www.exmaralda.org/en/), the Annotation Graph Toolkit (AGTK) (http://agtk.sourceforge.net/), the Express Scribe Transcription Software (http://www.nch.com.au/scribe/index.html) and InqScribe (https://www.inqscribe.com/). The Open Source Annotation Graph Toolkit (AGTK) uses annotation graphs as a formal framework for representing linguistic information, including annotations of linguistic behaviour from audio and video. Apart from its transcription and annotation module, the EXMARaLDA transcription and annotation tool includes a module for managing corpora as well as a module for computer-assisted querying of transcription, annotation and metadata. InqScribe is a less complex tool and is considered ideal for new users. In addition to transcription, the InqScribe transcription tool can also be used for creating subtitles for videos. One of the features of Google Live Transcribe is that it can caption real-time spoken words in over 70 languages and dialects and is available on 1.8 billion Android devices.[5]

Specialized tools for text analytics such as SentimentBuilder (http://sentimentbuilder.com/) can analyse unstructured texts such as emails, reviews or chat data with a Natural Language Processing module.

Some transcription tools are available as open source platforms. For example, the open source Oral History Metadata Synchronizer (OHMS) (http://www.oralhistoryonline.org) assists in relating textual search terms at word-level to the corresponding moment in the recorded data. The application produces a time-correlated transcript or indexed interview. Furthermore, there are online platforms such as "Recogito 2" (http://recogito.pelagios.org/) that allow collaborative document annotation, including images and the presentation of source materials and research results as Open Data (http://recogito.pelagios.org/). Online open source platforms are also available for learning languages and cultures, including

[5]https://venturebeat com.cdn.ampproject.org/c/s/venturebeat.com/2019/08/16/google-open-sources-live-transcribes-speech-engine/amp/ by Emil Protalinski.

ancient languages and literary heritage and culture. The Alpheios project allows the user to view a definition and morphological analysis of words chosen from web pages with Latin, Ancient Greek or Arabic content (http://alpheios.net/).

Example 2.4.1

Example of transcription tool with annotation of linguistic and paralinguistic information: The Anvil Video Annotation Research Tool (http://www.anvil-software.org).


Other tools, such as DM (Digital Mappaemundi), allow the annotation of texts, images and manuscripts. This facility allows users to create digitized resources from material such as manuscripts and photographs and to mark and comment selected fragments or to mark any indications of relations between texts or fragments of interest (https://schoenberg institute.org/dm-tools-for-digital-annotation-and-linking/). Available tools can create resources from scanned paper documents with the use of an optical character recognition engine. For example, the ABBYY FineReader creates editable and searchable electronic files from scanned paper documents, PDFs and digital photographs (https://www.abbyy.com/finereader/). Another example is the online work environment eLaborate (http://elaborate.huygens.knaw.nl/), where results of uploaded scanned, transcribed and annotated text are freely available to all users.

All the above-presented types of transcription and annotation tools can be employed in the processing, analysis and evaluation of spoken political and journalistic texts. Table 2.4.1 summarizes the common types of available transcription tools and the types (or level) of annotation provided.

**Table 2.4.1 Common types of available transcription tools and type of annotation**

| Common types of available transcription tools | Type (or level) of annotation |
|---|---|
| transcription software, online transcription tools / services<br><br>open source platforms<br><br>transcription software of Speech Recognition Software<br><br>transcription tools for creating subtitles for videos<br><br>annotation of images and manuscripts | basic text transcription<br><br>linguistic information<br><br>prosody<br><br>gestures |

## 2.5. Strategy for Processing Spoken Political and Journalistic Texts

### 2.5.1 Requirements, Parameters and Processing Levels

The nature and complexity of spoken political and journalistic texts does not always enable all features of their content and structure to be clearly outlined and, subsequently, processed.

Furthermore, the correct and efficient processing of spoken political and journalistic texts is also dependent on two basic parameters.

The first parameter is the above-described multi-domain content of most spoken political and journalistic texts. The multi-domain content is related both to issues concerning expressions and terminology and to issues concerning stylistic and pragmatic aspects of the spoken texts:

terminology may range from typical terms encountered in politics to specialized scientific terminology; stylistic features may be related to sociocultural or psychological factors; and pragmatic aspects may include argumentation and the expression of opinion.

The second parameter is the efficient use of existing tools and resources for machine translation and transcription.

Due to the features of the multi-domain content of most spoken political and journalistic texts, it is considered difficult for a machine translation application to produce output without any mistranslations. Spoken political and journalistic texts are considered complex texts, and the complexity factor increases when native speakers of two or multiple languages take part in a discussion or interview, especially if the transcribed spoken text is subjected to machine translation. Furthermore, in some cases, the processing of specific political and journalistic texts may be considered confidential, and any documents or other files related to the output may be treated as confidential data.

An additional issue in the correct and efficient processing of spoken political and journalistic texts is the fact that there are limitations in respect of the use of potentially useful resources such as spoken corpora. These limitations are due to the fact that existing resources with spoken corpora do not cover the entire range and particularities of spoken political and journalistic texts, especially in cases where international speakers have their own "variety" of speech and behaviour, which is often a combination of linguistic and paralinguistic features from their native language and the foreign language spoken.

- **A General Framework both for Human Analysts-Evaluators/Translators and for NLP Applications**

In relation to the above, a general framework is proposed for both the correct and efficient processing and the analysis of various types of spoken political and journalistic texts. Specifically, the general framework allows the implementation of a variety of automatic or interactive procedures, depending on text content and issues encountered. Spoken journalistic texts may be processed with respect to three different levels, corresponding to different types of complexity and levels of analysis, processing stages and respective processing speeds.

The proposed general framework concerns both the correct and efficient analysis and/or processing (A) by human evaluators, analysts or translators and (B) by Natural Language Processing (NLP) applications,

including Machine Translation and Data Mining (Opinion Mining/Sentiment Analysis).

The three levels of analysis and processing target the visibility and "processability" of the various aspects of complex information content in spoken political and journalistic texts.

The three levels of analysis are connected to distinctive types of tags serving annotation purposes and/or activating specific processes with the proposed strategies. Furthermore, the distinctive types of tags may also be used for the enrichment of corpora and empirical data for training language models – statistical models and neural networks.

### 2.5.2 Analysis Levels and Compatibility with Existing Tools and Systems

- **Analysis Levels**

As previously described, the information content of the spoken political and journalistic texts spans various levels of analysis that may be classified into three general categories.

The first level (Level 1) concerns the analysis of terminology and ambiguous phrases where an automatic or interactive "safety-mode" ambiguity resolution is proposed, corresponding to "fast-track"/ "stepping-stone" options used in Spoken Dialogue Systems (Lewis, 2009; Valavani et al., 2015). The "fast-track"/"safety-mode" options are proposed for quick, batch processing and analytical, interactive processing respectively. These processes may be enabled or disabled, depending on the type of processing and level of analysis activated.

Complex semantic information and connotative features are processed at Level 2. Some types of information can be automatically signalized while for other types of information, interactive processes are activated.

The processing of prosodic and paralinguistic features corresponds to Level 3, with the signalization of elements containing information such as prosodic emphasis and gestures. Depending on the type of Automatic Speech Recognition (ASR) System used, prosodic processing may either be automatic or interactive.

The proposed interactive annotation tool for "Terms, Complex Information and Prosody" (referred to as TCIP-Annotation) (Figure 2.5.1) concerns the annotation of the features of Levels 1, 2 and 3 namely the analysis of terminology and ambiguous phrases (Level 1), complex semantic information and indicators of Cognitive Bias (Level 2), and prosodic and paralinguistic features (Level 3).

For the annotation of terminology and ambiguous phrases requiring additional processing (Level 1), the "[SMI]" tag is proposed (presented in Chapter 3), connected to the respective strategies applied and linked to the respective resources. For the activation of the respective strategy/process, the tag may function as a "button". For terminology occurring in spoken journalistic texts, the inserted tags assist the additional processing or direct (machine) translation, including subtitles.

For the annotation of complex semantic information and connotative features (Level 2), the "[IMPL]" (text level)/"[impl]" (word level) tag is proposed, corresponding to processes of the respective implemented strategies. Strategies concerning the processing of features corresponding to Level 2 are presented in Chapter 4.

For the annotation of prosodic and paralinguistic features (Level 3), a set of specialized tags is proposed. The tags for prosodic and paralinguistic features also assist in processes detecting (complex) implied information (including degree of neutrality and Cognitive Bias). This information contributes to the evaluation of speaker behaviour. Strategies concerning the processing of features corresponding to Level 3 are presented in Chapter 5.

- **Compatibility with existing Tools and Systems**

One of the intended functions of the proposed TCIP-Annotation tool is its use as an additional annotation option to existing transcription tools and speech processing applications. The tag types concern text output generated by a Speech Recognition (ASR) module for pre-processing/post-processing, providing options for evaluation, (machine) translation or other processes, including Data Mining applications. The annotation can be run as an additional process or with a possible integration (as an upgrade) into existing tools and systems.

Additionally, the proposed TCIP-Annotation option (Figure 2.5.1), designed to operate on transcribed spoken texts, may also be adapted for the processing of spoken messages, especially for the indication of socio-cultural factors in incoming texts.

It should be stressed that the proposed Interactive Annotation for Terms, Complex Information and Prosody (TCIP-Annotation) serves as a general framework for integration into existing tools and systems and for adaptation according to the parameters and requirements of the natural language(s) and/or the application(s) concerned. The general framework property of the proposed TCIP-Annotation facilitates its use beyond a