

# The Context of War and Cognitive Bias: An Interactive Approach in Accessing Relations of Attitude, Behavior and Events in Ancient Texts and Online News

Christina Alexandris<sup>1,3</sup>(⊠), Jiali Du<sup>2</sup>, and Vasilios Floros<sup>1,3</sup>

<sup>1</sup> National and Kapodistrian University of Athens, Athens, Greece

calexandris@gs.uoa.gr, florosbas2002@yahoo.gr

<sup>2</sup> Guangdong University of Foreign Studies, Guangzhou, China

dujiali68@126.com, 201310039@oamail.gdufs.edu.cn

<sup>3</sup> European Communication Institute (ECI), Danube University Krems, Austria and National Technical University of Athens, Athens, Greece

Abstract. Access to knowledge from ancient classical texts for a comparison and understanding of current events and situations poses challenges, especially to non-experts, especially if information searched concerns behaviors, attitude and mentality linked to war and not facts, events and names. Geopolitical and diplomatic information concerning behavior and intentions is also connected to the challenge of precision, correction and capturing subtle details. These resources may be a valuable -yet often obscure- source of information to a broader User group, requiring expertise, language skills and a remarkable period of time to access and to evaluate these resources in order to combine and compare information with the current state-of-affairs. Easy access to the Classical Texts and the display of detailed and/or specific information in a user-friendly interaction is the main target of the designed user-interface and partially implemented applications. The approach integrates expert-knowledge and user requirements and also manages Cognitive Bias. The designed user-interface and partially implemented applications by-pass Cognitive Bias but also take advantage of specific types of Cognitive Bias.

Keywords: Cognitive Bias · Interface Design · Seed Ontology

# 1 Comparing Online News and Ancient Texts

# **1.1** Challenges in Accessing Information for the Understanding, Comparison and Evaluation of Attitude, Behavior and Events

The analysis and study of classical ancient historical texts concerning politicalgeopolitical and military information is a common practice for non-academic professionals of the field of Geopolitics and Diplomacy, such as military personnel, diplomats and journalists. In the present approach and applications, expert knowledge is targeted to be integrated in an attempt to utilize resources from World History in evaluating the current state of affairs and in decision-making. In particular, the targets involve knowledge extracted from ancient texts in respect to politics, diplomacy, human nature and attitudes linked to government, war, conflicts, internal politics and relations with other countries and powers.

Classical ancient historical texts provide (1) information as reference work, which is compared to the current state-of-affairs and (2) text structure allowing a better analysis and organization of the text content for the creation of possible models.

Non-experts and many categories of professionals alike may wish to access knowledge from ancient classical texts for a comparison and understanding of currents events and situations. Information from spoken political and journalistic texts and online written political and journalistic texts may be compared and evaluated in respect to knowledge and information to classical texts. However, unlike most types of language resources, both the linguistic features of the ancient texts concerned and the related knowledge and expertise do not facilitate access to a broader, international public and non-experts. Recent research and accomplishments in the field of Digital Humanities [18, 22] may provide a large variety of resources. However, accessing in-depth dimensions of their information content to a broader and international public remains a challenge.

Geopolitical and diplomatic information constitutes a type of information that is not easily processed with standard Information Extraction practices, since it does not concern mere facts, but behavior and intentions. In this case, the type of information content concerned does not allow a direct training and implementation of "off-the-shelf ontologies" [6] without the requirement of hand-labeled training data, as in the case of medical data [6]. Geopolitical and diplomatic information concerning behavior and intentions is also connected with the challenge of precision, correction and capturing subtle details. Furthermore, experts and professionals in the field of geopolitical and diplomatic information benefit from sources containing experience from the Past, describing geopolitical states-of-affairs, rhetoric and diplomacy. These resources may be a valuable yet often obscure source of information to a broader User group, requiring (a) expertise, (b) a remarkable period of time to access and to evaluate these resources in order to combine and compare information with the current state-of-affairs and (c) language skills, in many cases.

We note that these types of texts often are characterized by a particular structure of text and information and use of vocabulary. This means that the employment of expert knowledge in the analysis of the text structure and content is necessary. This is of essential importance since the text content is not in a (dated) modern language such as English or German as is the case of other classical references in the domain of War (for example, Carl Philipp Gottfried (or Gottlieb) von Clausewitz: "Vom Kriege" (About War), Alfred Thayer Mahan "The Influence of Sea Power upon History": 1660–1783).

The main challenge concerned is the process of guiding non-expert users and expert users alike in searching the respective information in the ancient classical texts. The content, language and structures of these text types requires a specialized customization of search techniques in addition to standard Information Extraction practices. This is especially necessary in the case of accessing information in regard to behaviors, attitude and mentality linked to war and not only facts, events and names. Precision and expert knowledge are factors contributing to an interactive approach versus a fully automatic approach in accessing and extracting complex information such as behavior and intentions resulting to decisions and events in History and Geopolitics. Employing an interactive approach in this case is similar to the practice of employing interactive approaches in interactive versus fully automatic approaches in Machine Translation – especially in cases where precision and completeness in the transfer of information content are of crucial importance.

As an example of ancient texts of World History, the "Peloponnesian War" of Thucydides (Ancient Greek) is taught in military academies, such as West Point. The present application concerns Ancient Greek historic texts, specifically, the "Peloponnesian War" of Thucydides, however, the general modelling approach used can be a starting point for possible adaptations to the specifications of other (ancient) texts, also in other languages.

Sun Tzu's "The Art of War" (Ancient Chinese) was written about 515 BC to 512 BC during the Spring and Autumn Period. Sun Tzu summarized the theories and principles of war and discussed the nature of war, the planning of war, danger in war, the preparations for war, strategic means in war, material supply in war, the general deployment in war, the analysis on situations in war, military virtue of an army in war, the use of the special battle in war (fire attack and espionage warfare) - among others.

#### 1.2 Design Specifications and Cognitive Bias

For the proposed interface, first of all, the user's search strategy should address the question whether information from classical references is relevant and useful in respect to understanding and/or analyzing current events and state-of-affairs. This information should be directly accessed in the respective texts and passages. Furthermore, specialized and/or detailed information should be made available, according to the type of classical text accessed. These specifications are summarized by the following questions: "If?", "Where?" and "How?".

The general approach for the modeling and implementations presented here are compatible with practices in spoken dialog systems allowing the options of both fast and slower-paced interaction [14]. The "If?" question corresponds to quick interaction whereas the "Where?" and "How?" questions correspond to intermediate and slower-paced levels of interaction speed [14] respectively.

For the modelling of the interface the way the information is presented to the user (Presentation) as well as the quality of information presented (Content) are of equal significance.

The present designed interface and partially implemented applications are based on modelling the knowledge provided by the (1) combination of existing language resources and expert knowledge contained in them, with (2) a strategy similar to practices employed in editors for Controlled Languages. This strategy is applied because the historic texts and translations concerned are resources with sublanguage-specific sub-domains and text-specific features presenting expert knowledge, formulated in the writer's or translator's style. Furthermore, the sub-domain of War and Geopolitics - Diplomacy in these resources allows the creation of ontologies and the formalization of entity relationships from text content and text structure such as Source-Outcome/Cause-Result relationships.

# 2 User Requirements and Cognitive Bias

### 2.1 Cognitive Bias

The designed user-interface and partially implemented applications for accessing knowledge from ancient classical texts targets to by-pass Cognitive Bias - but also to take advantage of specific types of Cognitive Bias. In particular, types of Cognitive Bias such as "Anchoring Bias", "Confirmation Bias" and "Bandwagon Effects" [4] are avoided whereas types of Cognitive Bias such as "Availability Bias", and "Framing Effects" [4] are used to the advantage of the interface creation and application implementation. The main target is to allow easy access to the Classical Texts and display detailed and/or specific information in a user-friendly interaction. The Cognitive Biases in the user's search are listed as following, according to Azzopardi, 2021:

"Availability Bias leads people to overestimate the likelihood of an answer or stance based on how easily it can be retrieved and recalled." [4] "Framing Effects occur when people make different decisions given the same information because of how the information has been presented" [4] "Anchoring Bias stems from people's tendencies to focus too much on the first piece of information learnt, or observed (even if that information is not relevant or correct)". [4] "Confirmation Bias stems from people's tendency to prefer confirmatory information, where they will discount information that does not conform to their existing beliefs." [4] "Bandwagon Effects occur when people take on a similar opinion or point of view because other people voice that opinion or point of view. Researchers have been concerned that search engines may be influencing people's opinions, either by presenting confirmatory information reinforcing people's existing beliefs [...], or by presenting information to sway their decisions through exposure effects (dubbed the Search Engine Manipulation Effect (SEME)." [4] Searchers rated articles as more useful if they were easier to read and understand" [4].

#### 2.2 Expert Users and Cognitive Bias

For professionals, precision and correctness are of crucial importance in information searched in ancient and historical texts (Requirement A), constituting a resource of expert knowledge from lessons learnt from the Past. If the information from these resources is to be compared with the current spoken journalistic and political texts, especially for decision-making, quick access to the requested content is a desired feature (Requirement B). Additionally, User requirements regarding the content of the information to be extracted were formulated with the aid of a questionnaire made available to prospective Users, especially journalists and military personnel.

Questionnaire –based User Requirements confirm that information from the Past can be relevant to the understanding of the current-state-of affairs, with the following topics consisting typical examples: In particular, Users strongly agreed with the following factors playing a crucial role in understanding Cause-Result relations in current affairs, directly related to geopolitical and diplomatic information from the "Peloponnesian War" of Thucydides: "Expert"-Users believed that the following applied in most cases: "Pressure from Allies is always a major factor", "Citizens' emotions are an unpredictable factor in decision-making", "Personality of leader is crucial in success of strategy", "Even today, war may be lost due to bad advisors". Users believed that the following applied in some cases: "Events may be explained by seemly irrelevant incidents" and "Unpredictable behavior of Allies may be due to factors related to domestic politics".

We note that the a-priori knowledge of expert-users is linked to word-entities and expressions such as "unpredictable", "emotions" and "domestic politics". However, a-priori knowledge, on the other hand, may also result to Confirmation Bias, where confirmatory information may be preferred, according to one's knowledge and experience. In other words, expert-users are not exempt from Cognitive Bias.

In order for the practices of professionals to be simulated by the implemented application, the nature and complexity of the information to be extracted requires the integration and formalization of expert knowledge as a starting point of analysis and investigation (Requirement C).

#### 2.3 Non-expert Users and Cognitive Bias

In contrast to expert-users, non-expert users may not be aware of the types of information content in the ancient texts. Therefore, types of Cognitive Bias such as Availability Bias and Anchoring Bias related to the accessibility and completeness of information are characteristic examples of Cognitive Bias that may be related to non-expert-users. Non-expert users may not always be able to evaluate the quality of precision and correctness of information searched in ancient and historical texts (Requirement A), however, as in all applications, precision is a essential requirement - for all users.

Quick access to the requested content is a desired feature (Requirement B) and may be a requirement of particular importance to non-expert users. Specifically, expertusers have, by default, due to their interests and expertise, a higher level of interest and engagement in the use of the application accessing information content of the ancient texts. Non-expert users have, by default, a lower level of interest and engagement in the use of the application and any errors and delays may discourage the users from any further interaction.

For non-expert users, the integration of expert knowledge (Requirement C) for accessing information in the ancient texts concerned involves the creation of a user-friendly interface, allowing easy interaction and formulation of queries.

Since the nature of the information to be extracted is not always easy to formulate as a query, because - in contrast to most applications - behavior, attitude and politics in general are concerned, the User's query is designed to be assisted by the sublanguage of the application. The sublanguage specifications and resulting ontologies function both for the search and extraction of information and for assisting User queries.

The interface targets to adapt the user's queries to the "world" of the ancient text. Specialized modules targeting to process distinctive linguistic features of the ancient texts are a necessary function. The same query word(s) can be related to different contexts and types of information in the Classical texts. The User cannot know or foresee the possible contexts and variations in information content of the query word.

In the case of ancient Classical Texts, information is often presented in a different way than in most (international) online texts. This results to non-expert and/or international users facing difficulties in accessing the information. In addition, non-expert and/or international users may not have a complete overview of the type of information available in the texts. Therefore, the targeted basic functions of the application interface are:

- to allow direct access to information not easily extracted
- to connect spoken texts from the live stream of current event to their "echoes" of related information in the resources concerned, in the present case, the resources from the ancient Past and
- to provide the necessary resources for understanding and providing possible clusters and associations of complex information concerning behavior and diplomacy.

# 2.4 Interaction and Specialized Functions

As described above, the design and creation of the interface focuses both on (a) the way the information is presented to the user (Presentation) and (b) the quality of information presented (Content)- in accordance with Requirements A, B and C.

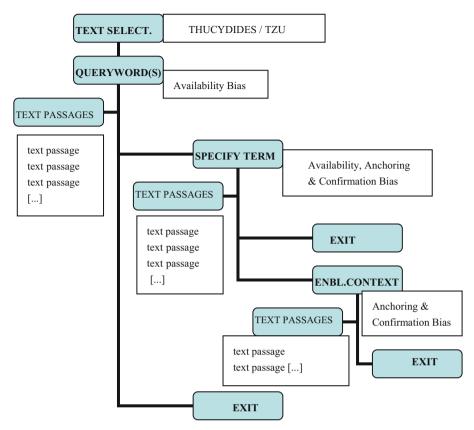
In respect to the presentation of the information, the interaction concerns the following basic steps: (1) the selection of the ancient text to be processed (TEXT-SELECT), (2) the insertion of the query word(s) (QUERY-WORD), (3) the refinement of the search (SPECIFY-TERM) and (4) the viewing of the specialized query word(s) options/choices displayed in the interface (SPECIFY-TERM results). Additional, optional steps are (5) the display of specialized and/or detailed information upon request (ENBL.CONTEXT) and (6) the extension and upgrading of the ontology used in the search process extension and upgrading of the search ontology - recommended for expert users (UPGRADE-ONTOLOGY).

The insertion of the query word(s) (2) and the refinement of the search (3) address the questions of "If?" and "Where?" and the specialized and/or detailed information (5) additionally addresses the question of "How?". The "Availability Bias" and "Framing Effects" Cognitive Biases [4] are used to the advantage for the modelling of the interface design and interaction.

Regarding the content of the information, the following three features are targeted to ensure the quality of the information presented: Choice of translation(s), text-specific and language-specific parameters of content and linguistic features, sublanguage-specific seed domain ontologies [8, 17]. These features are integrated in the design and implemented modules of the application for the avoidance of Cognitive Bias such as "Anchoring Bias", "Confirmation Bias" and "Bandwagon Effects" [4].

The presentation (Presentation) and (Content) of information are linked to the activation of the "Specify Term" and "Enable Content" processes in the interface. The activation of the "Specify Term" process with the respective button in the interface assists the user's query by guiding the search and providing possible options in respect to the information content of the word. In other words, if the word(s) inserted from the online journalistic text (QUERY-WORD) cannot be directly linked to corresponding passages in the Ancient Text, the activation of the "Specify Term" process refines the search.

After the words and expressions from user queries are matched to the content of the ancient Classical Text and the respective passages are displayed on the interface, the user may choose to activate the "Enable Context" button for the optional display of specialized and/or detailed information, according to the Classical Text selected. The optional fifth



**Fig. 1.** Overall outline and framework of the basic steps of the interaction and the optional "Enable Context" (ENBL.CONTEXT) function.

step of the interaction activated by the "Enable Context" process displays the word(s) selected within their possible contexts of specialized and detailed information.

For example, in the case of the Thucydides' Peloponnesian War, with the activation of the "Enable Context" button, a connection of the query word(s) with a sequence of "Cause-Result" and other types of relations related to the behavior of politicians and citizens is displayed in the respective passages. In the case of the "Art-of-War", the activation of the "Enable Context" button displays passages with characteristic forms of structure and content expressing significant information such as repetitions or other types of content characteristic of the ancient text, such as the skills and qualities of military leaders.

The "Update Ontology" button (optional step 6) allows the extension and upgrading of the ontology used in the "Specify Term" search process.

# **3** Content and the "Enable Context" Function

#### 3.1 The "Enable Context" Function: Corpora and Translations

The integration of expert knowledge and the choice of multiple corpora and translations are the necessary condition and requirement for the development and implementation of the "Specify Term" and "Enable Content" functions.

The choice of multiple corpora and translations allowing a broader range of options and information minimizes the possibility of Cognitive Bias, in particular, Availability Bias and especially Anchoring Bias and Confirmation Bias [4].

Translations, often bound by language-specific factors, often may not reflect the style of author and the "patterns" of the original text – an important feature in the text structure and content – and sometimes may not convey subtle but essential types of information [7]. The latter is important for the full transfer of the original information, while the feature of text structure contributes to detecting and extracting information. Ideally, translations of the original ancient text, along with additional "assistive" translations by scholars and experts whose native tongue is closely related to the language of the original text. This allows a larger set of structural and other linguistic similarities closer to the original text: Essential/subtle information may also be contained in the morpho-syntactic structure of words (or characters, in languages such as Chinese).

To maximize the coverage and precision of the information extracted, in previous research [16] multiple resources are used both in widely spoken languages such as English and with the above-stated "assistive" translations that can be processed with easily accessible and non-specialized online Machine Translation systems such as Google Translate. The resource for the English translation employed here is the MIT Classics Archive - Internet Classics Archive of the Massachusetts Institute of Technology (http://classics.mit.edu/Thucydides/pelopwar.mb.txt - Thucydides' Peloponnesian War, translated by Richard Crawley. J.M. Dent and Co., London 1903) [5].

The assistive "Katharevousa" translation here is the translation of the "Peloponnesian War" (in Katharevousa Greek - a "compromise" between Ancient Greek and Modern Greek mainly in formal texts and official documents, especially before the 1980's) by the prominent Greek statesman and political leader Eleftherios Venizelos (1864-1936). It was published in 1940 in the University of Oxford, after his death, also provided online (Centre for the Greek language: Portal for the Greek Language: www.greeklanguage.gr, E.Venizelos Translation [1940] 1960 [21]). The translation is very close to the original Ancient Greek text, however, it explicitly presents most of the information implied by pronouns and other forms of anaphora and context-dependent expressions in the original Ancient Greek text, facilitating the direct access to the text content with the use of the sublanguage-specific keywords. Furthermore, in this translation, an increasing number of causal relations is visible with pointers such as "because", which might not be available in an original English translation [2]. In order to be processed by Google Translate in English, the translation in "Katharevousa" Greek can be submitted to partial editing [2]. As observed from the evaluation of the translations from previous research [2], Google Translate could successfully handle the partially processed Katharevousa Greek text (Assistive Translation). The following example illustrates the additional information (in

brackets) from the Assistive Translation, as well as its similarity to the original ancient text (We note that the Athenians and Lacedaemonians (Spartans) were the superpowers of the time):

English translation (for Queries) (MIT Classics Archive): The Mantineans and their allies were the first to come over [become allies with] through fear of the Lacedaemonians. [Because] Having taken advantage of the war against Athens to reduce a large part of Arcadia into subjection, they thought that Lacedaemon would not leave them undisturbed in their conquests, now that she had leisure to interfere, and consequently gladly turned to a powerful city like Argos, the historical enemy of the Lacedaemonians, and a sister democracy.

Assistive Translation (for Search and Extraction): [5.29.1] Πρώτοι οι Μαντινείς και οι σύμμαχοί των προσεχώρησαν εις την συμμαχίαν ταύτην, εκ φόβου των Λακεδαιμονίων. Διότι, διαρκούντος ακόμη του προς τους Αθηναίους πολέμου, οι Μαντινείς είχαν υποτάξει μέρος της Αρκαδίας, και ενόμιζαν, ότι οι Λακεδαιμόνιοι δεν θα τους επέτρεπαν να διατηρήσουν την επ' αυτού κυριαρχίαν, ήδη οπότε αι χείρες των ήσαν ελεύθεραι. Ωστε προθύμως εστράφησαν προς το Άργος θεωρούντες αυτό πόλιν ισχυράν, και ανέκαθεν αντίπαλον των Λακεδαιμονίων, και επί πλέον δημοκρατουμένην, όπως και αυτοί.

Original Ancient Text: [5.29.1] Μαντινῆς δ' αὐτοῖς καὶ οἱ ξύμμαχοι αὐτῶν πρῶτοι προσεχώρησαν, δεδιότες τοὺς Λακεδαιμονίους. τοῖς γὰρ Μαντινεῦσι μέρος τι τῆς ᾿Αρκαδίας κατέστραπτο ὑπήκοον ἔτι τοῦ πρὸς Ἀθηναίους πολέμου ὄντος, καὶ ἐνόμιζον οὐ περιόψεσθαι σφᾶς τοὺς Λακεδαιμονίους ἄρχειν, ἐπειδὴ καὶ σχολὴν ἦγον· ὥστε ἄσμενοι πρὸς τοὺς Ἀργείους ἐτράποντο, πόλιν τε μεγάλην νομίζοντες καὶ Λακεδαιμονίοις αἰεὶ διάφορον, δημοκρατουμένην τε ὥσπερ καὶ αὐτοί.

We note that respective texts related to events from current spoken journalistic or political texts from which queries may be formulated were not presented here for reasons of political correctness.

A similar observation - in respect to the translation proximity to the ancient textis also observed in "The Art of War" by Sun Tzu. As quoted from Zheng (2019): "Lin Wusun is a Chinese scholar who is 19 years older than Roger T. Ames, so he has the advantage of understanding the original text". These differences are illustrated in the following example:

[24]: 静以幽, 正以治. This term is from chapter eleven, in which the whole sentence is: "将军之事, 静以幽, 正以治". Ames's translation: As for the urgent business of the commander: He is calm and remote, correct and disciplined. Lin's translation: It is the responsibility of the commander to be calm and inscrutable, to be impartial and strict in enforcing discipline [24].

### 3.2 The "Enable Context" Function: Implemented Modules and Parameters

**Parameters for the "Peloponnesian War" by Thucydides.** In previous research [1, 2], for the extraction of Source-Outcome/Cause-Result relations from the "Peloponnesian War" of Thucydides, a sublanguage-based approach was employed, based on the structure and linguistic features of the source text (Ancient Greek) and the linguistically related "Assistive" ("buffer") translation (Katharevousa Greek).

Standard Information Extraction techniques are based on the universal or textdependent (syntax) [15, 22, 23] logical relations between entities - facts, names, objects, actions as concepts, whether text-dependent or text-independent [3, 10, 11]. However, information related to mentality, intentions, beliefs and emotions as well as socio-cultural factors for the presentation and presentation of Source-Outcome/Cause-Result relations is not easily processed with standard Information Extraction techniques. This is due to the fact that the above described type of information is not easily analyzed and categorized in sublanguage-independent detectable and extractable entity groups and patterns of sequences of words and entities. Although typical practices in Digital Humanities provide a necessary basis for any forms of Information Extraction, here, the employment of a customized approach is necessary.

Additionally, precision and correctness are here a basic requirement, as in the case of technical texts, where practices of a traditional in-depth analysis are used to create Controlled Languages. The strategy employed is a sublanguage-based formalization of ontologies in the vocabulary and sentence structure, typical in Controlled Languages [12, 13] which were originally based on features of technical texts and extended to other task-oriented domains and applications. To conform to the requirements of precision and correctness but also to achieve speed for an easy access to the requested information, a strategy with features and practices of Controlled Languages is employed, such as controlling input in relation to a restricted set of words and processing predefined types of sentence structure related to respective types of content.

The implemented application is based on three types of ontologies. These ontologies are used for the extraction of the requested information in the text passages presented: The Topic-Keyword Ontology (the actual word-entries from the online political and journalistic texts as user-input -extendable by the User); The assistive Query Ontology (visible to the User); The Search Ontology (extracting the passages from multiple [other] corpora- primarily the "Assistive" Translation).

We note that these ontologies used in previous research presented here [1, 2, 16] correspond to the proposed Seed-Domain ontologies depicted in Sect. 4, namely: The Topic-Keyword Ontology corresponds to the proposed general search ontology ("Start-Up" ontology, Sect. 4) and the Query Ontology and Search Ontology both correspond to the proposed singular "Search Term" ontology presented here in Sect. 4. These modifications target to simplify the search process. However, for analysis purposes, we present the original ontology types used in the implemented specialized application.

The Topic-Keyword Ontology and the sublanguage-specific Query Ontology are combined (TQ) and used to assist the User's query as a singular search list and to refine the User' search. The Search Ontology operates in resources consisting translations from ancient texts (MIT Classics Archive -Crawley Translation [5]) in combination with the "Assistive" translation in languages closer to the original ancient texts (Portal for the Greek Language – Venizelos Translation [21]). The Search Ontology functioning as a search and extraction tool is based on the Source-Outcome/Cause-Result relationships explaining politics, diplomacy and geopolitical relations from the "Peloponnesian War" of Thucydides.

The sublanguage-specific Query Ontology [1, 2] is in English and in the language of the "Assistive" Translation (Katharevousa Greek). It is used to assist the User's query

and to refine the User's search and is based on keywords clustered around three basic concepts. These concepts extend the formalization of the sublanguage of "Diplomacy" from previous studies [2]: State (for example: neutrality" or "disadvantage"); Action (for example: "response"- "reaction"- "answer" or "accept" and "rejection"); Result (for example: "gain"- "benefit"- "profit" or "loss"). Furthermore, the Query Ontology contains an additional small set of words with sublanguage-specific tags such as "Athenians-[Superpower]", to assist Users queries (currently approximately 280 words).

The specialized Search Ontology performs the actual search in the translation close to the ancient text. The strategy employed for extracting the requested information from passages in the "Assistive" translation is based on: (1) the recognition of a defined set of conjunctions (CONJ) and (2) the recognition of a set of words concerning intention and behavior, annotated as "Intention-Behavior"- IB words (verbs and participles). Multiple IB words contained in passages extracted can be related to a singular query containing keywords from the Topic-Keyword Ontology and the Query Ontology (TQ-Keyword):

## Query: [TQ-Keyword(s)] IB < CONJ > IB [TQ-Keyword(s)] [2, 16].

The IB words occur "before" and "after" the conjunction (CONJ). The text containing the IB word(s) before the conjunction CONJ expresses the "Result (Outcome)" relation and the text containing the IB word(s) after the conjunction CONJ expresses the "Cause (Source)" relation. However, for some types of conjunctions, the reverse order applies [2, 16]. The order and type of "Cause (Source)" and "Result (Outcome)" is dependent on the type of conjunction concerned. This type of order is defined according to the information structure in the Assistive Translation, which allows a strict formalization of information content based on syntactic structure similar to formalizations for creating Controlled Languages. This is the basis on which the Cause-Result relations are extracted.

The group of specified conjunctions describing causal relations contains expressions such as "because" and "due to" ("διότι", "επειδή", "άλλωστε", "δια το", "δηλαδή", "ένεκα", "ένεκεν", "ώστε").

Relations between topics may concern "IB verbs" of the following types:

"Feeling-Intention-Attitude" type (what was believed, what was felt, what was intended, what attitude prevailed -Int-Intention) (for example: "were intended to" ("διατεθειμένοι"), "ignored", "were ignorant about" ("ηγνόουν"), "expected", "calculated", "took into account" - "υπελόγιζαν");

"Speech-Behavior" type - Sp-Speech (what was said - for example: "asked", "demanded" ("εξήτουν"), "convinced" ("πείσουν"), "supported", "backed" - "υπεστήριζε");

"Benign-Malignant Behavior" type (actual behavior -Bh-Behavior) (for example: "secured" (in context of negotiation) (" $\epsilon \xi \alpha \sigma \varphi \alpha \lambda (\sigma \alpha \varsigma")$ . [2, 16].

In the following example (implementation in JAVA [16]), the passages contain Cause-Result relations related to the keywords "subjects (of superpowers)" from the Topic-Keyword Ontology paired with "revolt" and "carried away" from the Query Ontology (TQ). A query concerning the possibility of a revolution by people controlled by a superpower ("subjects (of superpowers)" "revolt") is refined and assisted with the aid of keywords from the Topic-Keyword Ontology and the Query Ontology. The search and extraction is performed by the Search Ontology (IB verbs and CONJ), extracting one or multiple passages containing the keywords from the Topic-Keyword Ontology:

Query Content: [subjects (of superpowers), revolt (TQ)] Search Ontology match: (IB-Int: showed desire) <CONJ:because> (IB-Sp: admit)

The extracted passages of the matches in the text are presented to the User (The Eighth Book, Chapter XXIV, Nineteenth and Twentieth Years of the War - Revolt of Ionia - Intervention of Persia - The War in Ionia). The additional information from the Assistive Translation (Katharevousa Greek text) is depicted in square brackets, as demonstrated in the following example:

English Translation: But above all, the subjects of the Athenians showed a readiness to revolt [against rule] even beyond their ability, [because] judging the circumstances with [carried away by] [revolutionary] passion, and refusing even to hear of the Athenians being able to last out the coming summer.

Assistive Translation: Before: CONJ ("διότι")-IB: "εκτιμώμεναι":[Αλλ' οι, υπήκοοι προ πάντων των Αθηναίων εδείκνυαν μεγάλην επιθυμίαν όπως αποτινάξουν την κυριαρχίαν των και αν ακόμη αι δυνάμεις των ορθώς εκτιμώμεναι δεν ήσαν επαρκείς εις τούτο]

After: IB: "παραδεχθούν": [διότι εις τας κρίσεις των παρεσύροντο από τον επαναστατικόν οργασμόν, και δεν ήθελαν να παραδεχθούν καν ότι οι Αθηναίοι ήτο ενδεχόμενον να ανθέξουν κατά το προσεχές θέρος].

We note that keywords in the Topic-Keyword Ontology and the Query Ontology (TQ) may also be subjected to Machine Translation. In the previous approaches concerned [1, 2], this included Universal Words, with the use of the Universal Networking Language (UNL - www.undl.org) originally created for processing UN documents in languages as diverse as English, Hindi and Chinese [20].

**Parameters for "The Art of War" by Sun Tzu.** In the case of "The Art of War" by Sun Tzu, we note the characteristic use (and repetition) of verbal anaphora in the text content and structure [19] and particular types of military terms [24].

For the detection of repetitions, indicating content of emphasized significance, the respective process for the detection of this specialized and detailed information can be activated by the "Enable Context" button.

Example [24]: "十六字诀" This term refers to "上兵伐谋, 其次伐交, 其次伐兵, 其下攻城", which is from chapter three of Sun Tzu.

Ames's translation: The best military policy is to attack strategies; the next to attack alliances; the next to attack soldiers; and the worst to assault walled cities.

Lin's translation: The best policy in war is to thwart the enemy's strategy. The second best is to disrupt his alliances through diplomatic means. The third best is to attack his army in the field. The worst policy of all is to attack walled cities.

Example [19]: 是故军无辎重则亡, 无粮食则亡, 无委积则亡. (军争 第七)

Pinyin: Shigu jun wu zizhong ze wang, wu liangshi ze wang, wu weiji ze wang. (Chapter 7 Manoeuvring)

Version (1). An army without its baggage train is lost; without provisions it is lost; without bases of supply it is lost. (L. Giles)

217

Version (2). For this reason, if an army is without its equipment and stores, it will perish; if it is without provisions, it will perish; if it is without material support it will perish. (Roger Ames)

Diverse meanings of a seemingly singular term are another characteristic feature of "The Art of War" by Sun Tzu. For example, according to the description in "The Art of War", there are five types of spies in war, and  $\sum |\vec{n}|$  (făn jiàn) is the third one. The discussion about the military terminologies of spies in "The Art of War" is as follows:

Example of diverse terms from "The Art of War" by Sun Tzu:

- (1) The first kind of spy is called 因间 (yīn jiàn). Sun Tzu said"因间者 (yīn jiàn zhě), 因其乡人而用之 (yīn qí xiāngrén ér yòng zhī)". That means 因间 is a special spy whom you can make work for you because he is your fellow countryman. Therefore, the military term of 因间 (yīn jiàn) can be translated into "a fellow countryman spy".
- (2) The second kind of spy is called 内间 (nèijiàn). "内间者 (nèijiàn zhě), 因其官人 而用之 (yīn qí guānrén ér yòng zhī)". That means 内间 is a special spy whom you can make work for you because he is an enemy government official. Therefore, the military term of 内间 (nèijiàn) can be translated into "an enemy government official spy".
- (3) The third kind of spy is called 反间 (fǎnjiàn). "反间者 (fǎnjiàn zhě), 因其敌间而 用之 (yīn qí díjiān ér yòng zhī)". That means 反间 is a special defecting spy whom you can make work for you because he is an enemy spy. Therefore, the military term of 反间 (fǎnjiàn) can be translated into "an enemy's converted spy".
- (4) The forth kind of spy is called 死间 (sǐjiàn). "死间者 (sǐjiàn zhě), 为诳事于外 (wèi kuáng shì yú wài), 令吾间知之 (lìng wújiān zhīzhī), 而传于敌间也 (ér chuán yú díjiān yě)". That means 死间 is a betraying spy who is doomed to die. To deceive the outside world, we intentionally make certain information collected and disseminated to the enemy by the spy who betrayed us. The military term of 死间 (sǐjiàn) can be translated into "a doomed spy".
- (5) The fifth kind of spy is called 生间 (shēngjiàn). "生间者 (shēngjiàn zhě), 反报也 (fǎn bào yě)". That means 生间 is a surviving spy who can come back alive to report. The military term of 生间 (shēngjiàn) can be translated into "a surviving spy".

The implementation of modules concerning the processing and evaluation of the content of political and journalistic texts from international English-speaking news networks can be adapted to the content type and linguistic features selected English translations of the "The Art of War" by Sun Tzu. These modules implemented in previous research [2, 16] involve the signalization of occurrences of word repetitions in sentences and paragraphs of political and journalistic texts from international news networks [16]. The implemented modules also involve the signalization of particular word classes and their percentages with the aid of the Stanford Log-Linear Part-of-Speech Tagger [2, 16, 25]. Unlike the content of political and journalistic texts from international news networks, where mostly nouns, proper nouns, adjectives and adverbs were selected and processed [2, 16], in the case of "The Art of War by Sun Tzu, verbs and verbal anaphora play a significant role. In contrast to proper names and nouns, verbs and verbal anaphora constitute word types that are less commonly selected by application users in Information Extraction strategies and other forms of search mechanisms. In other words, there is a lexical category bias in respect to verbs and verbal anaphora [9]. However, in the case of "The Art of War by Sun Tzu, these less commonly sought word types are observed to be links to essential information.

The different word types are detectable with a POS Tagger. The respective words and word categories may constitute a small set of entries in a specially created lexicon or may be retrieved from existing databases or WordNets. In this case, sublanguage-specific Seed Domain ontologies.

# 4 Presentation: Modelling the "Specify Term" Function

### 4.1 Modelling Domain-Specific Seed Ontologies

As described above, the Interface is designed to provide possible options to the User in order to assist the query. These options are both in respect to the information content of the word ("Specify Term") and its possible contexts ("Enable Context" button).

In the previously mentioned "Specify Term" function, the User chooses a word from an online written or spoken text and enters it into the interface. The User selects the type of Classical Text concerned and then activates the "Specify Term" button. The function of query assistance minimizes the possibility of Cognitive Bias, in particular, Availability Bias [4].

In the first step of the interaction of the "Specify Term" function, namely the "Select query word(s)" sub-task, individual keywords of User queries in form of free input are recognized by the application. Words and expressions from the free input of User queries can be directly matched to the content of the ancient Classical Text. The user can directly proceed with search and extraction in the respective passages displayed in the interface. If there is no match, the user proceeds to the next step of the interaction of the "Specify Term" function, namely to "Refine search words" sub-task (sublanguage-specific) ("Specify Term" function - Message: No matches found. Proceed with Search anyway? OR Refine Search). In this case, keywords from sublanguage-specific ontologies appear as an option (selected from the menu or as a pop-up window) to assist the User's query. In particular, the User's query is assisted by ontologies presented as a singular search list in the interface of the application ("Specify Term" function as an interactive Controlled Language editor for the management of input for texts to be processed (for example, max. 20 English words of average size).

These ontologies constitute a sublanguage-specific domain ontology [22] functioning as a prior knowledge [8, 17], which can be extended and adapted, if necessary, for example, for term clustering with seed knowledge-based LDA models [8]. These ontologies constitute hand-labeled training data for further training and implementation [17]. As described above, we note that the nature of the information content concerned does not allow a direct training and implementation of ontologies without the requirement of hand-labeled training data, as in other types of information and related applications [6]. There is a general search ontology (Start-Up ontology) for both types of Ancient Texts. To facilitate User's queries, the extendable Start-Up ontology contains a "start-up" set of predefined general sublanguage expressions such as "war", "allies" and "rebellion" of currently approximately 100 words (Start-Up).

There is a specialized ontology for each type of text, namely the Thucydides Search Term Ontology and the Sun Tzu Search Term Ontology, which are connected to the respective "Enable Content" modules specializing in the information content of the type of Ancient Text. One ontology is based on terms from Thucydides' Peloponnesian War and the other ontology is designed to be based on the terms of "The Art of War" by Sun Tzu. Both ontologies are extendable.

With the "Update Ontology" function, the (expert) user may choose to update and/or extend the general search ontology (Start-Up ontology) or one of the specialized Search Term Ontologies, the Thucydides Search Term Ontology or the Sun Tzu Search Term Ontology. The sublanguage-based formalization of ontologies in the vocabulary and sentence structure allows the use of keywords, a feature typical of Spoken Dialog Systems, where speed is of crucial importance [14]. The Topic-Keyword Ontology is extendable. There is the option for its extension by expert users ("Update Ontology" function Menu: Save Query) (not presented in Fig. 1).

Typical examples from the Seed-Domain ontology of Thucydides' Peloponnesian War are general search ontology (Start-Up ontology) terms "war", "allies" and "rebellion" and text and sublanguage-specific terms connected to the respective nodes in the Search Term Ontology, namely "State", "Action" and "Result". Examples of sublanguage-specific terms connected to the "Action" node are "response", "reaction", "answer"," accept" and "rejection". Examples of sublanguage-specific terms connected to the "Action" node are "response", "reaction", "answer"," accept" and "rejection". Examples of sublanguage-specific terms connected to the "Result" node are "gain", "benefit", "profit" and "loss" (Fig 2).

Typical examples from the Seed-Domain ontology of "The Art of War" by Sun Tzu include the above-described general search ontology (Start-Up ontology) terms "war" "allies" and "rebellion" and text and sublanguage-specific terms concerning "Military policy". They are connected to the Search Term Ontology subset of the "The Art of War" Seed-Domain ontology. The Search Term Ontology contains sublanguagespecific verbs and the respective nodes namely "Strategy", "Battlefield-Tactical" and "Objects". Examples of sublanguage-specific terms connected to the "Strategy" node are "military policy", "strategies", "alliances", "enemy", "diplomatic means", "soldiers". Examples of sublanguage-specific terms connected to the "Battlefield-Tactical" node are " enemy field", "(walled) cities", "bases of supply", also "soldiers". Examples of sublanguage-specific terms connected to the "Objects" node are "baggage train", "provisions", "equipment", "stores", "material support" (Fig. 3).

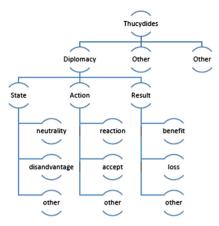


Fig. 2. Overview of Thucydides "Search Term" Ontology (Seed Domain Ontology).

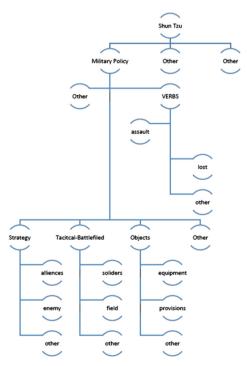


Fig. 3. Overview of Sun Tzu "Search Term" Ontology (Seed Domain Ontology).

# 5 Conclusions and Further Research

Information from spoken political and journalistic texts and online written political and journalistic texts is targeted to be linked to knowledge and information to classical texts for its comparison and evaluation. However, access to knowledge from ancient

classical texts for a comparison and understanding of currents events and situations poses challenges, especially to non-experts. This process can become of increased complexity especially if information is searched regarding behaviors, attitude and mentality linked to war and not facts, events and names. Easy access to the Classical Texts and the display of detailed and/or specific information in a user-friendly interaction is the main target of the designed user-interface and partially implemented applications. The approach integrates expert-knowledge and user requirements and also manages Cognitive Bias. The designed user-interface and partially implemented applications by-passes Cognitive Bias but also takes advantage of specific types of Cognitive Bias.

The upgrading and updating of the existing ontologies by expert-users is expected to play a key-role in the full implementation and overall improvement and upgrading of the application and interface. The target of the application and interface is its function as a "two-ended" collaborative search interface, with the user on the one end and the expert/expert-user on the other end, updating /upgrading the ontology.

A possible future step in the implementation of the application is the generation of models created from the relations between the words processed in the "Enable Context" functions. The strategy used in the present application can also function as a corpus builder, as a training platform and as starting point for further adaptations and additional goals. However, for any further developments full implementation of all the interface functions and extensive evaluations are necessary.

Acknowledgements. In loving memory of our co-author Vasilios Floros who passed away before the completion of the present research paper.

# References

- Alexandris, C., Mylonakis, K., Tassis, S., Nottas, M., Cambourakis, G.: Implementing a platform for complex information processing from written and spoken journalistic data. In: Kurosu, M. (eds.) HCI 2017. LNCS, vol. 10271, pp. 549–558. Springer, Cham (2017). https:// doi.org/10.1007/978-3-319-58071-5\_41
- Alexandris, C.: Accessing cause-result relation and diplomatic information in ancient "journalistic" texts with universal words. In: Kurosu, M. (eds.) HCI 2014. LNCS, vol. 8511, pp. 351–361. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07230-2\_34
- Angeli, G., Premkumar, M.J., Manning, C.: Leveraging linguistic structure for open domain information extraction. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Stroudsbourg, PA, USA, pp. 344–354. Association for Computational Linguistics ACL (2015)
- Azzopardi, L.: Cognitive Biases in search: a review and reflection of cognitive biases in information retrieval. In: Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2021), New York, NY, USA. ACM (2021). https://doi.org/ 10.1145/3406522.344602
- 5. Crawley, R.: Thucydides' Peloponnesian War, translated by Richard Crawley. J.M. Dent and Co., London (1903)
- Fries, J.A., et al.: Ontology-driven weak supervision for clinical entity classification in electronic health records. Nat. Commun. 12 (2021). https://doi.org/10.1038/s41467-021-223 28-4

- 7. Hatim, B.: Communication Across Cultures: Translation Theory and Contrastive Text Linguistics. University of Exeter Press, Exeter (1997)
- 8. Huang, H., Harzallah, M., Guillet, F., Xu, Z.: Core-concept-seeded LDA for ontology learning. Procedia Comput. Sci. **192**, 222–231 (2021). https://doi.org/10.1016/j.procs.2021.08.023
- 9. Jia, H., Liang, J.: Lexical category bias across interpreting types: implications for synergy between cognitive constraints and language representations. Lingua **239** (2020)
- Jurafsky, D., Martin, J.H.: Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 2nd edn. Prentice Hall series in Artificial Intelligence, Pearson Education, Upper Saddle River (2008)
- Jurafsky, D., Martin, J.H.: Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 3rd edn. (2022). Draft: https://web.stanford.edu/~jurafsky/slp3/ed3book\_jan122022.pdf
- Kuhn, T.: A survey and classification of controlled natural languages. Comput. Linguist. 40(1), 121–170 (2014)
- Lehrndorfer A.: Kontrolliertes Deutsch: Linguistische und Sprachpsychologische Leitlinien f
  ür eine (maschniell) kontrollierte Sprache in der technischen Dokumentation. T
  übingen, Narr (1996)
- Lewis, J.R.: Introduction to Practical Speech User Interface Design for Interactive Voice Response Applications. IBM Software Group, USA, Tutorial T09 presented at HCI 2009 San Diego, CA, USA (2009)
- Mausam, N.V., Schmitz, M., Bart, R., Soderland, S., Etzioni, O.: Open language learning for information extraction. In: Proceedings of EMNLP-CoNLL 2012 -Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Stroudsbourg, PA, USA, pp. 523–534. Association for Computational Linguistics ACL (2012)
- 16. Mylonakis, C.: Processing and extracting complex information in written and spoken journalistic data and historic (ancient) texts. Masters thesis, Department of Informatics and Telecommunications, National University of Athens, Greece (2016)
- Rawsthorne, H.M., Abadie, N., Kergosien, E., Duchêne, C., Saux, E.: ATONTE: towards a new methodology for seed ontology development from texts and experts. In: Corcho, O., Hollink, L., Kutz, O., Troquard, N., Ekaputra F.J. (eds.) EKAW 2022, Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management, Bolzano, Italy (2022). hal-03794323
- Suissa, O., Elmalech, A., Zhitomirsky-Geffet, M.: Text analysis using deep neural networks in digital humanities and information science. J. Assoc. Inf. Sci. Technol. (JASIST) 73, 268–287 (2022). https://doi.org/10.1002/asi.24544
- 19. Tao, Y.: Translation of verbal anaphora in "the art of war." Theory Pract. Lang. Stud. **3**(6), 1040–1044 (2013). https://doi.org/10.4304/tpls.3.6.1040-1044
- Uchida, H., Zhu, M., Della Senta, T.: Universal Networking Language. The UNDL Foundation, Tokyo, Japan (2005)
- Venizelos, E.: Thoukudidou Istoriai: Kata Metaphrasin Eleutheriou Benizelou. [Thucydides' History: translated by Eleftherios Venizelos]. In: Caclamanos, D. (ed.) Two Vols. Oxford University Press, Oxford (1940)
- Zhang, B., Kumpulainen, S., Keskustalo, H.: A review on information retrieval in the historical digital humanities domain. In: Proceedings of the Asian Conference on Arts & Humanities 2021 ACAH2021, Tokyo, Japan (2021). https://doi.org/10.22492/issn.2186-229X.2021.1
- Zhao, G., Zhang, X.: Domain-specific ontology concept extraction and hierarchy extension. In: Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval (NLPIR 2018), Bangkok, Thailand, pp. 60–64. ACM, New York (2018). https://doi.org/10.1145/3278293.3278302

- Zheng, J.: A comparative study on English translations of military terms in Sun Tzu: "the art of war". Theory Pract. Lang. Stud. 9(5), 537–544 (2019). https://doi.org/10.17507/tpls.090 5.07
- 25. The Stanford Log-linear Part-Of-Speech Tagger. https://nlp.stanford.edu/software/tagger. shtml. Accessed 6 Nov 2022