

# Implementing a Platform for Complex Information Processing from Written and Spoken Journalistic Data

Christina Alexandris<sup>1</sup>(✉), Konstantinos Mylonakis<sup>1</sup>, Stefanos Tassis<sup>1</sup>,  
Mario Nottas<sup>2</sup>, and George Cambourakis<sup>3</sup>

<sup>1</sup> National and Kapodistrian University of Athens, Athens, Greece  
calexandris@gs.uoa.gr, {kmylo, stassel}@di.uoa.gr

<sup>2</sup> Danube University Krems – National Technical University of Athens,  
Athens, Greece  
manottas@ilsp.gr

<sup>3</sup> National Technical University of Athens, Athens, Greece  
gcamb@cs.ntua.gr

**Abstract.** The implemented platform targets to assist in information processing and decision-making processes concerning online written and transcribed spoken political and journalistic texts. Two separate functions are linked, one evaluating the degree of implied information and connotative features and one allowing access and comparison to related content such as explanatory information and cause-result relations in historic (ancient) “journalistic” texts.

**Keywords:** Databases · Spoken texts · Connotative features · Ontologies

## 1 Interaction and Design

The present platform is intended to assist in information processing and decision-making processes concerning online written and transcribed spoken political and journalistic texts such as interviews, live conversations in the Media and discussions in Parliament. The platform links two separate functions with the possibility to (A) provide the User-Journalist with the largest possible percentage of the points in the texts signaling information with implied information and connotative features (Function A - “TextTone”) [2, 12] and to (B) allow access and comparison to related content such as explanatory information and cause-result relations in historic (ancient) “journalistic” texts (Function B - “Echo”) [1, 12].

The platform makes use of existing online tools and applications, such as Google Translate, and includes the activation of two databases for each separate function. Implementation is in JAVA [12]. The design and implementation is based on data and observations provided by professional journalists, Program M.A in Quality Journalism and Digital Technologies, Danube University at Krems Austria, Athena - Research and Innovation Center in Information, Communication and Knowledge Technologies, Athens - Institution of Promotion of Journalism Ath.Vas. Botsi, Athens.

The browsed and viewed journalistic texts can be subjected to an automatic evaluation of the percentage of non-neutral content and can be directly linked to related information from sources typically constituting reference material that cannot be easily accessed, such as ancient and historical texts. In other words, the current journalistic or political texts can be evaluated in the light of the degree and percentage of non-neutral content and in relation to similar- not identical- situations and lessons learnt from the Past.

From the main menu of the platform, the User selects the type of function to be activated. Function A (TextTone) requests an input of an online journalistic text or transcribed spoken text and generates all elements constituting “marked” information with implied information and connotative features. The TextTone function is a fully automatic process. Function B (Echo), a user-interactive process, operates with an input of selected keywords from current journalistic texts generating passages from the historic texts explaining chronological and “cause-result” relations in politics and diplomacy from the Past, allowing a comparison with events from the Present.

## **2 Evaluating Degree of (Non) Neutral Tone in Online Journalistic Texts: The “TextTone” Function**

The “TextTone” function concerns the signalization of all “connotatively marked” words and expressions in written and transcribed spoken journalistic texts. The function is based on the flouting of Grice’s Cooperativity Principle (Grice 1989, Hatim 1997) in Pragmatics theory, especially in regard to the violation of the Maxims of Quality and Quantity [2, 6, 7], since the procedure involves the differentiation between “superfluous” and “necessary” elements in a journalistic text, beyond the absolutely necessary “Who/What-When-Where-(How)” framework expressed by the “necessary” elements [2]. The “TextTone” function operates on a database and generates the percentage of words related to non-neutral content.

### **2.1 The “TextTone” Function Database**

“Connotatively marked” words and expressions whose semantic content is related to connotatively emotionally and socio-culturally “marked” elements may be grouped into a finite set based on word type, word stems or suffix type, namely at word level or at the morphological level [2, 12]. Signalization of word-classes such as adjectives and adverbials is implemented with the Stanford Log-Linear Part-of-Speech Tagger [15]. Additionally, recognition on a word-stem or a suffix basis involves the detection of types of verbs, related to specific semantic features (also accessible with Wordnets and/or Selectional Restrictions [2]).

The word groups concerned are the grammatical categories of (1) adjectives and (2) adverbials, containing semantic features related to (i) descriptive features (ii) mode (iii) malignant/benign action or (iv) emotional/ethical gravity [2, 12]. Word groups with Implied Connotative Features involve specific categories of (3) verbs (or nominalizations of verbs) containing semantic features (including implied connotations in

**Table 1.** Text with high score in “non-neutral” content, signalized marked features as output.

<a href="http://edition.cnn.com/2016/04/04/opinions/wisconsin-humble-political-celebrities-opinion-borsuk/index.html">http://edition.cnn.com/2016/04/04/opinions/wisconsin-humble-political-celebrities-opinion-borsuk/index.html</a>	52 matches
humble/JJ big/JJ big/JJ little/JJ humility/NN surprising/JJ loaded/VBN big/JJ like/VBP pretty/RB bit/NN humility/NN humiliation/NN remarkable/JJ best/JJS big/JJ much/JJ likely/JJ just/RB wondered/VBD bit/NN soundly/RB just/RB much/JJ interesting/JJ much/RB much/RB humbling/JJ humiliating/JJ sharply/RB just/RB bluntly/RB much/JJ imagine/VB big/JJ dicey/JJ think/VB great/JJ splintering/VBG just/RB bad/JJ humility/NN humbled/VBN big/JJ best/JJS seems/VBZ bit/NN humble/JJ humility/NN amazing/JJ big/JJ big/JJ	19% Adjectives: 8% Adverbs: 6% Other: 5%
soundly/RB sharply/RB frequently/RB bluntly/RB Literally/RB originally/RB lately/RB belatedly/RB necessarily/RB only/RB	
have/VB: 3 Donald/NNP: 2 Trump/NNP: 2 I/PRP: 2 was/VBD: 2 he/PRP: 2 he/PRP: 3 Hillary/NNP: 2 Clinton/NNP: 2 Sanders/NNP: 2 Sanders/NNP: 2 Clinton/NNP: 2 done/VBN: 2 better/JJR: 2 has/VBZ: 2 much/RB: 2 race/NN: 2 him/PRP: 2 Republican/JJ: 2 would/MD: 2 his/PRP\$: 2 would/MD: 2 you/PRP: 2 ‘ve/VBP: 2 that/WDT: 2	Word repetitions in same sentence

language use) related to (i) mode (ii) malignant/benign action or (iii) emotional/ethical gravity, as well as (4) nouns with suffixes producing diminutives, derivational suffixes resulting to a (ii) verbalization, (iii) an adjectivization or (iii) an additional nominalization of proper nouns [2, 12]. For journalistic texts, percentages over 18% (Table 1) in “connotatively marked” words signalize a high score of “non-neutral” content. A low score in “non-neutral” content is below 10%. (Table 2).

**Table 2.** Text with medium to high score in “non-neutral” content, signalized marked features as output.

<a href="http://www.bbc.com/news/election-us-2016-35962179">http://www.bbc.com/news/election-us-2016-35962179</a>	38 matches
much/JJ sharp/JJ sick/JJ quickly/RB fine/JJ disappointing/JJ just/RB decimated/VBN heartily/RB booted/VBN proud/JJ think/VBP important/JJ think/VBP nervous/JJ think/VBP think/VBP sick/JJ negative/JJ surprisingly/RB pessimistic/JJ flaring/VBG pretty/RB relentless/JJ unfortunately/RB worse/JJR think/VBP glad/JJ stumping/VBG harsh/JJ negative/JJ flaring/VBG calm/VB deep/JJ calm/JJ good/JJ just/RB much/RB disastrous/JJ	15% Adjectives: 7% Adverbs: 5% Other: 3%
New/NNP: 2 contests/NNS: 2 he/PRP: 2 her/PRP\$: 2 campaign/NN: 2 manufacturing/VBG: 2 her/PRP\$: 2 who/WP: 2 Democratic/JJ: 2 people/NNS: 2 vote/NN: 2 we/PRP: 2 he/PRP: 2 state/NN: 2 their/PRP\$: 2 as/RB: 4 far/RB: 2 support/NN: 2 I/PRP: 5 can/MD: 2 talk/VB: 2 I/PRP: 2 their/PRP\$: 3 are/VBP: 3 he/PRP: 2 Sanders/NNP: 2 campaign/NN: 2 back/RB: 2	Word repetitions in same sentence

### 3 Current Journalistic Texts and “Echoes” of the Past

Current journalistic and political texts with a high percentage of non-neutral content often concern topics of high importance. These topic types may be registered and compared with similar topics linked to phenomena and situations of the Past for assisting decision-making, for example, in order to avoid conflict, subjection or war. However, information from the Past is often limited to experts and scholars, especially if a different language is concerned. Experts and professionals in the domain of political and journalistic texts compare and contrast events, policies and behavior of the Past to the state-of affairs in the Present.

#### 3.1 User Requirements

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: “Pressure from Allies is always a major factor” (1a), “Casus-Belli is characterized by an evident Cause-Result relation” (1b), “In geopolitical maps some features change, others remain the same” (1c).

Users believed that the following applied in most cases (2): “Citizens’ emotions are an unpredictable factor in decision-making” (2a), “Personality of leader is crucial in success of strategy” (2b), “Geopolitics is connected both to the Past and to long-term plans for the future” (2c), “Today, war is not very different: there are other means, plus the factor of globalization” (2d), “Even today, war may be lost due to bad advisors” (2e).

Users believed that the following applied in some cases (3): “Events may be explained by seemingly irrelevant incidents” (3a), “Unpredictable behavior of Allies may be due to factors related to domestic politics” (3b).

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 (USA). 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.

For professionals, details in historic texts are a must and a generalized type of comparability with information from the Past is often not sufficient: Precision and correctness are of crucial importance in information searched in ancient and historical texts (Requirement A), as 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. In accordance to the practices of professionals to be simulated by

the present 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).

### 3.2 Combining Online Journalistic Texts with “Echoes” of the Past

The basic functions of the application are to allow direct access to information not easily extracted and to connect spoken texts from the live stream of current events to their “echoes” of related information in the resources concerned, in the present case, the resources from the ancient Past. The sublanguage-based formalization of ontologies in the vocabulary and sentence structure allows the use of keywords, a feature typical of Dialog Systems, where speed is of crucial importance. A database containing keywords related to the domain of Diplomacy and Politics relates topics from the current-state-of-affairs to related passages from the ancient Past.

These keywords are related to predefined ontologies [2] to assist the User’s query (Interface Message: Use only nouns, verbs and adjectives), to facilitate search and to extract the requested information. The keyword ontology assisting the User’s query can be extended and upgraded (Interface Menu: Save Query). The platform integrates additional user-input, upgrading and updating existing ontologies generated from the interaction.

The sublanguage-specific ontology used to assist the User’s query and to refine the User’s search is based on keywords clustered around basic concepts related to the sublanguage of “Diplomacy”: (1) topic, (2) state, (3) action and (4) result, as an extension from previous studies [2, 5].

The concept of “state” (Category “state”) contains singular words or expressions such as “neutrality” or “disadvantage”. The concept of “actions” (Category “actions”) contains expressions such as “response” - “reaction” - “answer” or “accept” and “rejection”. The concept of “result” (Category “result”) contains expressions such as “gain” - “benefit” - “profit” or “loss”. The sublanguage-specific ontology may be referred to as the “Query Ontology”. Furthermore, the Query Ontology (Q-Ontology) contains an additional small set of words with sublanguage-specific tags, such as “Athenians-[Superpower]”, to assist Users queries.

For accessing information from ancient texts, keywords are subjected to Machine Translation (MT) prior to any further processing by the “Echo” function. Machine Translation may involve available online MT applications, such as Google Translate or special MT systems and databases, such as the Universal Networking Language (UNL) originally created for processing UN documents in languages as diverse as English, Hindi and Chinese [14].

## 4 Direct Access to Related Information from Ancient Texts: The “Echo” Function

The second function (“Echo”) concerning the processing of complex information in ancient texts intends to address queries regarding diplomatic and political problems, their resolution, correct or bad decisions, mistakes and socio-cultural phenomena

related to politics. Although the ancient texts concerned are the texts of Thucydides “Peloponnesian War”, these texts may be regarded as a starting-point for further adaptations and upgrading for processing ancient historical texts of other languages, with the integration of expert knowledge, as described in the following strategy, which re-introduces traditional ontology-based approaches and Controlled Languages.

#### 4.1 Strategy

The employed strategy is based on the link of the above presented sublanguage-related ontology with resources of expert knowledge. The ontology enables direct access to the translated ancient texts, an approach that does not employ standard Information Extraction techniques. In particular, the strategy avoids the process of adapting requested information – which is, in the present case, related to mentality, intentions, beliefs, emotions and socio-cultural factors- to practices based on the universal or text-dependent (syntax) logical relations between entities/facts categorized in sublanguage-independent detectable and extractable entity groups and patterns of sequences of words/entities [4, 8, 11]. On the other hand, to conform to the requirements of precision and correctness (Requirement A) but also to achieve speed for accessing requested information also from spoken texts (Requirement B), the strategy integrates practices typical in Controlled Languages [9, 10], by utilizing information contained in vocabulary as well as text and sentence structure. Specifically, a restricted set of words and predefined types of sentence structure related to respective types of content are processed. In the present case, texts facilitating such types of processing are translations very close to the original Ancient Greek text, explicitly presenting most of the information implied by pronouns and other forms of anaphora and context-dependent expressions in the original Ancient Greek text. In these texts, a large number of causal relations is visible with pointers [3] such as “due”, which might not be available in other translations [2].

The translations concerned are in formal Modern Greek or “Katharevousa”, a “compromise” between Ancient Greek and the Modern Greek, in particular, the translations by prominent Greek statesman and political leader Eleftherios Venizelos (1864–1936), 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](http://www.greeklanguage.gr), E. Venizelos Translation [1940] 1960) [13]. These translations combining both linguistic proximity to the original text and expert knowledge function as an Assistive-“Buffer” translation, connecting the User’s queries with keywords translated from English (or another language) to Greek and presenting respective passages from English translations (or in another language). Therefore, expert knowledge is integrated in the present strategy (Requirement C), often with additional features and information not always visible in translations in other languages, mainly due to linguistic parameters [1]. The example in Table 3 related to the keyword query “allies” and “change sides” 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).

**Table 3.** Example of query keywords (“allies”, “change sides”) and sample of related passages.

Query: “allies”, “change sides”
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 Μαντινεῖς και οι σύμμαχοί των προσεχώρησαν εις την συμμαχίαν ταύτην, εκ φόβου των Λακεδαιμονίων. Διότι, διαρκούντος ακόμη του προς τους Αθηναίους πολέμου, οι Μαντινεῖς είχαν υποτάξει μέρος της Αρκαδίας, και ενόμιζαν, ότι οι Λακεδαιμόνιοι δεν θα τους επέτρεπαν να διατηρήσουν την επ’ αυτού κυριαρχίαν, ήδη οπότε αι χεῖρες των ἦσαν ελεύθεραι. Ὡστε προθύμως εστράφησαν προς το Ἄργος θεωρούντες αὐτό πόλιν ισχυράν, και ἀνέκαθεν ἀντίπαλον των Λακεδαιμονίων, και ἐπὶ πλέον δημοκρατουμένην, ὅπως και αὐτοὶ
Original Ancient Text: [5.29.1] Μαντινῆς δ’ αὐτοῖς και οἱ ζῦμμαχοι αὐτῶν πρῶτοι προσεχώρησαν, δεδιότες τοὺς Λακεδαιμονίους. τοῖς γὰρ Μαντινεῦσι μέρος τι τῆς Ἀρκαδίας κατέστραπτο ὑπήκουον ἔτι τοῦ πρὸς Αθηναίους πολέμου ὄντος, και ἐνόμιζον οὐ περιόψεσθαι σφᾶς τοὺς Λακεδαιμονίους ἄρχειν, ἐπειδὴ και σχολὴν ἦγον· ὥστε ἄσμενοι πρὸς τοὺς Ἄργειους ἐτράποντο, πόλιν τε μεγάλην νομίζοντες και Λακεδαιμονίους αἰεὶ διάφορον, δημοκρατουμένην τε ὥσπερ και αὐτοὶ

## 4.2 The “Echo” Function Database

The extraction of 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).

The word groups in the “Echo” Function Database may be referred to as the “Search Ontology”. One or multiple IB words contained in passages extracted can be related to a singular query containing keywords from the above-mentioned keyword Query Ontology (Q- Keyword):

Query: [Q- Keyword(s)] IB <CONJ> IB [Q- Keyword(s)] [12].

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. 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, outlining the content of the relations.

The group of specified conjunctions describing causal relations contains expressions such as “because” and “due to” (“διότι”, “ἐπειδὴ”, “ἀλλωστε”, “δια το”, “δηλαδὴ”, “ἐνεκα”, “ἐνεκεν”, “ὥστε”).

**Table 4.** Query (Q-keywords: “allies”, “revolt”) IB words and other Q-keywords in match.

SPOKEN TEXT [Current Journalistic/Political Text] “allies, “revolt”
User QUERY: [Q- Keyword(s)] IB <CONJ> IB [Q- Keyword(s)]
[subjects (of superpowers), revolt (Q)]
(IB-Int: showed desire) <CONJ:because> (IB-Sp: admit)
[carried away (Q), passion (Q)]

Relations between topics may concern IB verbs: (I) of “Feeling-Intention-Attitude” type, what was believed, what was felt, what was intended, what attitude prevailed (Int-Intention), (II) of “Speech-Behavior” type (Sp-Speech), what was said and (III) of “Benign-Malignant Behavior” type, actual behavior (Bh-Behavior). The types of IB verbs are tagged (Int, Sp, Bh) for possible use in other applications.

Examples of the “Feeling-Intention-Attitude” type (Int) are verbs such as “were intended to” (“διατεθειμένοι”), “ignored”, “were ignorant about” (“ηγνόουν”), “expected”, “calculated”, “took into account” (“υπελόγιζαν”). Typical examples of the “Speech-Behavior” type (Sp) are the verbs “asked”, “demanded” (“εζήτησαν”), “convinced” (“πείσουν”), “supported”, “backed” (“υπεστήριξε”). An example of the “Benign-Malignant Behavior” type (Bh) is “secured” (in context of negotiation) (“εξασφαλίσαμε”).

In the following example (implementation in JAVA) [12], the passages contain Cause-Result relations related to the keywords “subjects (of superpowers)”, “revolt” and “carried away” from the Query Ontology (Q).

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 Query Ontology. Search and extraction is performed by the Search Ontology (IB verbs and CONJ), extracting one or multiple passages containing the keywords from the Query Ontology: (Table 4).

The extracted passages are presented to the User (The Eighth Book, Chapter XXI, 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 [12]: (Table 5).

**Table 5.** Example of an extracted passage for Q-keywords: “allies”, “revolt”.

<http://www.mikrosapoplous.gr/thucy/vivlia/vivlio8.htm>

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: “εκτιμώμενοι”: {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 οργανισμόν, και δεν ήθελαν να παραδεχθούν καν ότι οι Αθηναίοι ήτο ενδεχόμενον να ανθέξουν κατά το προσεχές θέρος}



## 5 Conclusions and Further Research

In the constant development of media, access to information expands “horizontally”, across various user groups, but gaining an in-depth insight with precision often remains a challenge. In addition, implied information and connotative features as well as benefits of past knowledge, which may be described as a “vertical”, in-depth dimension of information, are seldom exploited.

The present platform targets to facilitate the access to “vertical”, in-depth information for a broader user group. This type of information may be characterized as complex information, concerning mentality, intentions, beliefs, emotions and socio-cultural factors. Processing of complex information in online written and transcribed spoken journalistic and political texts, involves signaling implied information and connotative features by the TextTone function, as well as accessing and comparing related content and explanatory information in historic texts by the Echo function. Both functions target to assist in information processing and decision-making processes. In the Echo function, the nature and complexity of the information processed calls for the re-introduction and employment of traditional ontology-based strategies.

Both functions constitute a basis for upgrading and possible adaption to texts of other languages, if applicable. For the TextTone function, this depends on whether implied and connotative features are retrievable in the morphosyntactic or lexical level of linguistic analysis. For the Echo function, the extraction of the type of complex information concerned also depends on text structure and style of the ancient author. Further adaptation and implementation of both functions will provide an insight to possible additional parameters in the strategies presented.

## References

1. Alexandris, C.: Accessing Cause-Result relation and diplomatic information in ancient “Journalistic” texts with universal words. In: Kurosu, M. (ed.) HCI 2014. LNCS, vol. 8511, pp. 351–361. Springer, Cham (2014). doi:[10.1007/978-3-319-07230-2\\_34](https://doi.org/10.1007/978-3-319-07230-2_34)
2. Alexandris, C.: English, German and the international “Semi-professional” translator, a morphological approach to implied connotative features. *J. Lang. Translation* **11**(2), 7–46 (2010). Sejong University, Korea
3. Carlson, L., Marcu, D., Okurowski, M.E.: Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue, Eurospeech 2001, Denmark, September 2001
4. 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, Association for Computational Linguistics, pp. 344–354, Beijing, China, 26–31 July 2015
5. Crawley, R.: *Thucydides’ Peloponnesian War*. J.M. Dent and Co., London (1903). Translated by Richard Crawley
6. Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J. (eds.) *Syntax and Semantics*, vol. 3. Academic Press, New York (1975)

7. Hatim, B.: *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*. University of Exeter, Exeter (1997)
8. Jurafsky, D., Martin, J.: *Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall Series in Artificial Intelligence, 2nd edn. Pearson Education, Upper Saddle River (2008)
9. Kuhn, T.: A survey and classification of controlled natural languages. *Comput. Linguist.* **40**(1), 121–170 (2014)
10. Lehnrdorfer, A.: *Kontrolliertes Deutsch: Linguistische und Sprachpsychologische Leitlinien für eine (maschniell) kontrollierte Sprache in der technischen Dokumentation*. Narr, Tübingen (1996)
11. 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, Proceedings of the Conference, Association for Computational Linguistics*, pp. 523–534 (2012)
12. 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)
13. Venizelos, E.: *Thoukudidou Istoriai: Kata Metaphrasin Eleutheriou Benizelou*. [Thucydides' History: Translated by Eleftherios Venizelos], Caclamanos, D. (ed.), vol. 2. Oxford University Press, Oxford (1940)
14. Uchida, H., Zhu, M., Della Senta, T.: *Universal Networking Language*. The UNDL Foundation, Tokyo (2005)
15. The Stanford Log-linear Part-Of-Speech Tagger. <http://nlp.stanford.edu/software/tagger.shtml>