



Volume XXVI 2023

ISSUE no.2

MBNA Publishing House Constanta 2023



## Scientific Bulletin of Naval Academy

SBNA PAPER • OPEN ACCESS

### The use of semantic analysis in the identification of attacks based on social engineering techniques

To cite this article: R. Moinescu, C. Răcuciu, C.-I. Păiuș and D. Glăvan, Scientific Bulletin of Naval Academy, Vol. XXVI 2023, pg. 74-80.

Submitted: 22.03.2023

Revised: 30.07.2023

Accepted: 10.08.2023

Available online at [www.anmb.ro](http://www.anmb.ro)

ISSN: 2392-8956; ISSN-L: 1454-864X

doi: 10.21279/1454-864X-23-I2-008

SBNA© 2023. This work is licensed under the CC BY-NC-SA 4.0 License

# The use of semantic analysis in the identification of attacks based on social engineering techniques

Radu MOINESCU, Ciprian RĂCUCIU, Cătălin-Ionuț PĂIUȘ, Dragoș GLĂVAN

Military Technical Academy "*Ferdinand I*"  
radu.moinescu@gmail.com

**Abstract.** The need to detect attacks based on social engineering techniques has put researchers in the position of building different tools, observation grids or even theories able to capture different behaviors specific to simulated behavior. Although phishing is the main attack vector used in most cybersecurity offensive playbooks, the most effective attacks involve dialogue between the attacker and the target. This paper explores threat detection using semantic analysis performed with "Tropes" software.

## 1. Introduction

Social engineering is a process of influencing the behavior of a person, by manipulating his/her behavior, or by gaining and subsequently betraying his/her trust, in order to gain access to computer systems controlled by the person(s) concerned. Social engineering techniques differ depending on the medium used: email, phone, social media or even in person. The purpose of a social engineering attack is to build a relationship with the target, gain their trust, and induce the target to provide information that violates their organization's policy or personal security practices. This type of attack has grown rapidly in recent years, and for certain categories of targets, the poorly prepared, is the dominant technique. For this reason, it is considered that "man is the weakest link in a security system".

In essence, in order to protect ourselves against attacks based on social engineering techniques we need to answer the following questions:

- What information needs to be protected?
- Why should it be protected?
- What would be the consequences if the information is compromised?
- If during conversation, questions arise regarding that situation, how should one react so that it is not a direct refusal or uncivilized behavior, starting from the idea that it is possible that it is a coincidence that the person approached that topic?

## 2. Papers related to the analysis available in the public domain

LiarOrNot allows the identification of certain relevant micro-expressions from the perspective of emotions that a person experiences when interacting with someone else, betraying his intentions, feelings, which he currently thinks he masks quite well.[1] Paul Ekman, for example, after extensive research has determined that, despite the fact that people can display over 10,000 distinct facial expressions, using a series of facial muscles that extremely finely control the lips, mouth, nose, eyes, forehead, cheeks they can be recognized as conveying basic emotions such as sadness, happiness, anger, fear, surprise, disgust and contempt - by all people, regardless of their culture. Emotional expressions are the same in all cultures.[2] Moreover, most researchers agree that there is an innate ability to express joy, anger, sadness, surprise and despair.[3] The rules of externalization also carry a cultural imprint to

a certain extent: social rules may dictate that we diminish, exaggerate, completely hide or mask an emotion that we feel at one time or another. For example, when watching a movie about accidents, the Japanese had a much greater tendency to mask the negative emotion with a smile.[2]

The identification of the direction of the gaze and the analysis of this data in a first form can be used not only in marketing, but also as hypotheses in the evaluation of sincerity, using the theory of neuro-linguistic programming - more specifically we refer to the concept of status indicators, promoted by Richard Bandler, John Grinder.[5] Neuro-linguistic programming provides some crucial information that can give serious clues about what a person is thinking, whether they are lying or not, and in which sensory modality they prefer to express themselves.

The voice can be analyzed with vocal stress analysis software, also revealing emotional connotations that can be exploited. The software highlights any changes in tone that may be indicators of insincerity or stress. Computer Voice Stress Analyzer (CVSA), is an example of such a program.[6]

Through semantic analysis, in a complementary relationship with social network analysis, specialists can contribute, on the one hand, to the identification of the main risk elements that could affect a field and, on the other hand, to the development of additional tools/methods imperatively necessary for anticipatory/preventive actions. Also, through a multidisciplinary approach, based on scientifically validated methods, which resides in the permanent concern of the scientific community for objectivity, by using the software, analysts (from any field of activity, including economic, sociological, etc.) can acquire capabilities to answer a series of questions, depending on interest.

### **3. Semantic analysis of social engineering attacks using "Tropes" software**

#### *3.1 "Tropes" Software Description*

Developed by Pierre Molette and Agnès Landré, based on the work of Rodolphe Ghiglione, "Tropes" is a natural language processing and semantic classification software that guarantees the relevance and quality of text analysis. It also offers an extensive range of tools and instruments and semantic analysis, through which answers can be obtained regarding the content of a text, through a referential analysis, this fact also contributes to the explanation of some phrases in the field of linguistics.

By carrying out a classification at the semantic level, "Tropes" can outline specific semantic categories, which are interdependent with the interest of the beneficiaries of the final result; these are called the script, which includes parts of speech such as nouns, verbs, adjectives, connectors.

The usefulness of the "Tropes" software resides in its application, for example, in tracking the identification of the ways of manifestation, in semantic and semasiological terms, of the concepts transposed into the specialist's sphere of interest; thus, after the implementation of the materials required to be subjected to analysis in the "Tropes" software, it is possible to highlight the mode of exposure of the style/mode of expression used (descriptive, argumentative, reflexive, etc.), the typology of the context (dynamic/actional, static), the identification the most important characteristic parts of the text and the component episodes of the analysis, which may be conclusive for the intended objective.

As a linguistic unit of communication, the statement is in an interdependent relationship, in its realization, with the availability offered by the linguistic system of a language, an imperative fact necessary for the correct selection of software versions, depending on the language in which it is elaborate the material/text that will be implemented in the program (for example, an English text cannot be subjected to an analysis of the Romanian version of the software and vice versa, the results being completely inconclusive).[7]

At the same time, "Tropes" contributes to the organizational breakdown of the statement, as a discursive unit, which is directly dependent on the information transmitted and the situational context of communication, dividing it into structured versus unstructured statement; thus, the software highlights the unstructured utterances, represented by phonetic elements, identified by words or phrases, which communicate, most of the time, information regarding the speaker's reaction to an extralinguistic fact.

Last but not least, "Tropes" also help to highlight structured utterances, which can be subjected to analysis in the case of oral discourse, transcribed ad-literam, including possible elements of speaker hesitation.

In this sense, the software allows the identification of the most frequently used concepts, the totality of the relationships in the conceptual system being highlighted by the graphic representation of all categories of words used in the text, belonging to an extended range of parts of speech (verbs, adjectives, pronouns, numerals, connectors).

A particular case can be represented by the initiation of a complementary semantic analysis, in the context of the oral discourse, which bears the imprint of pauses, discontinuities in thought or expression, determined by the spontaneous nature of oral expression, marked by: repetitions or voluntary self-repetitions, the use of certain verbs or interjections, of some conjunctions, of some words associated with verbal tics (sometimes, verbs or adverbs of attitude appear in middle positions, so that they can turn into a strategy of masking/disguising hesitation); also strongly marked in oral communication are the ways of expressing imprecision, specific to the registers associated with orality: nouns (a bit, a dram, etc.), adverbs (somewhat), undecided adjectives (something, some), identified by the software "Tropes" analysis as elements of doubt.

In this case, the software helps to establish a psychological identity of the interlocutors, such as those who show reluctance to engage in communication, in the context of which they use many repetitions, a large number of words / marks of hesitation, have a thematic extent LOW.

### 3.2 Analysis and detection of phishing emails

It is fully known and accepted that the success of attacks based on social engineering techniques is due to the exploitation of human weaknesses such as: helpfulness, sexual attraction, convenience, compassion, gullibility, curiosity, fear, greed, respect for authority, sense of urgency, sympathy or vanity.

In this case, the use of the semantic analysis software "Tropes" has proven its applicability for the purpose of analyzing phishing e-mails that can draw a conclusion about the malicious people who use emotions against other people in order to make a quick emotional decision, without being passed through the filter of thought.

Thus, in the current case study, several such e-mails were subjected to analysis, in the form of a database consisting of passages from these materials within the semantic analysis software program "Tropes", which proposes semantic analysis tools, with the help of which answers to questions about the content of a text can be obtained, by means of a referential analysis.

It also helps to create its own semantic categories, customized according to the beneficiary's interest, which can be represented by nouns, verbs, adjectives, connectors. In addition, they can be extended with the help of software dictionaries, depending on the targeted analysis strategy. In order to process the texts subject to analysis, "Tropes" facilitated operations at the level of: delimiting phrases and sentences; solving the presence of ambiguities (depending on the appearance of some words in the text); identifying classes of equivalent words; detecting the characteristic parts of a context.

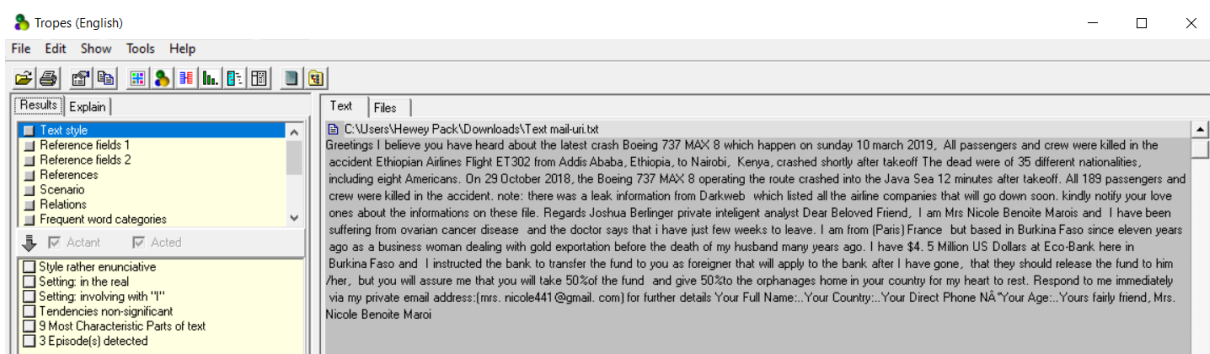


Fig.1. Integration of phishing e-mail in "Tropes" software

After integrating the material into the software program (Fig.1), it emphasized that the style/mode of expression used is enunciative; the context is dynamic/actional; 9 characteristic parts of the text and 3 episodes/component parts of the analysis were identified, however, partially conclusive for the objective targeted in the study.

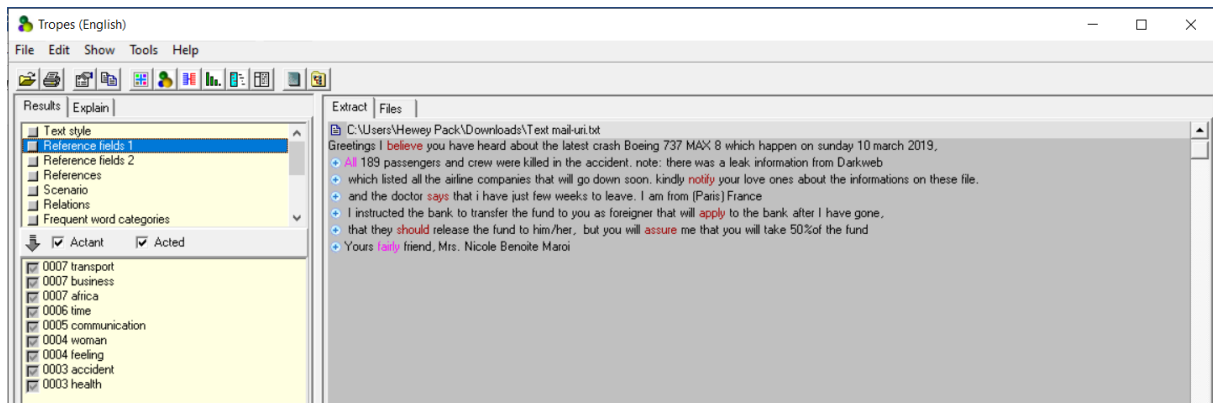


Fig.2. Semantic analysis of phishing e-mail in "Tropes" software

Thus, regarding the context under analysis, it was noted that it has the peculiar feature found in awareness articles developed by specialists, in order to bring attention to a field or topic, in this case, that of phishing (Fig.2).

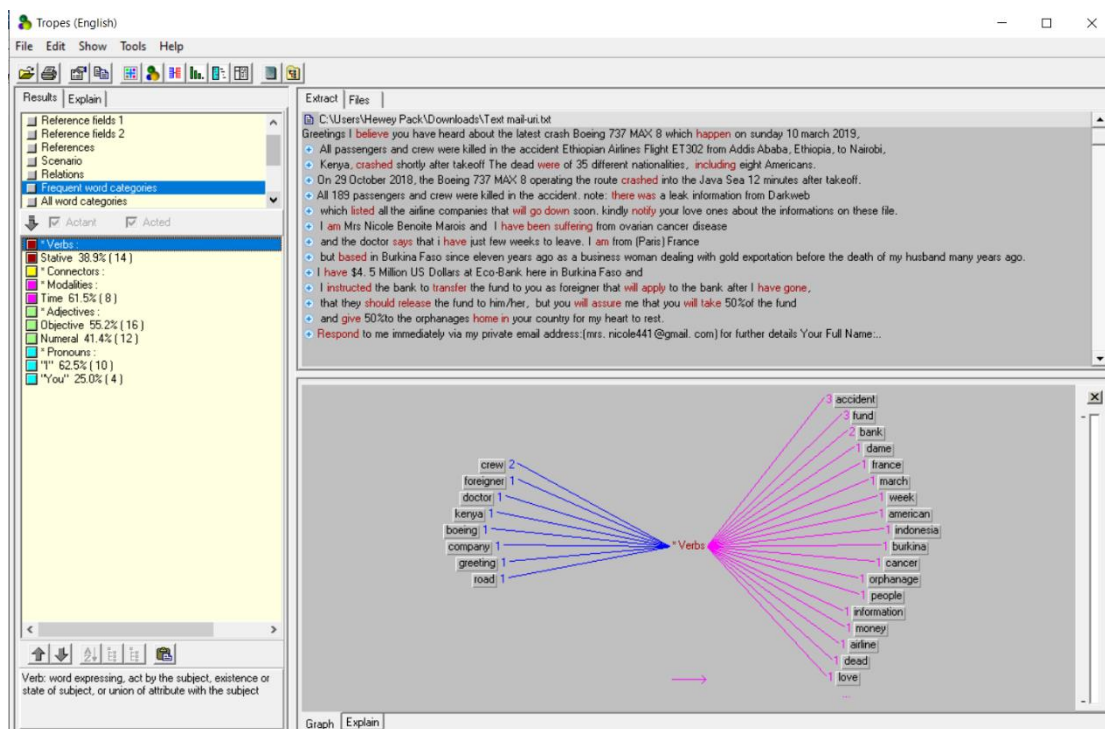


Fig.3. The relationship of verbs in the phishing e-mail

Thus, a multilingual terminology can be noted, reflected by the use of factive verbs, "I think", "collapsed", "suffered", "I was trained", "I assure you". In this sense, the class of factive verbs is included among the categories of the most frequently used verbs: "to be", "to have", "to do", used to determine emotions (Fig.3). Emotion, by its very nature, is the change in readiness for action to maintain or change the relationship with an object or event. Motivation or motivational change is one of the key aspects of

emotions. Even so, action follows only under certain conditions, including the presence and availability of an action repertoire, a balance of costs and benefits of action, and the presence of resources and motivation to consider costs and benefits. There are trade-offs between repertoire selection and cost-benefit considerations. The repertoire usually includes low-effort actions that greatly expand the influence of emotions on action.

As can be seen in the image below (Fig.3.), there are certain categories of verbs used to elicit or appeal to the recipient's emotion.

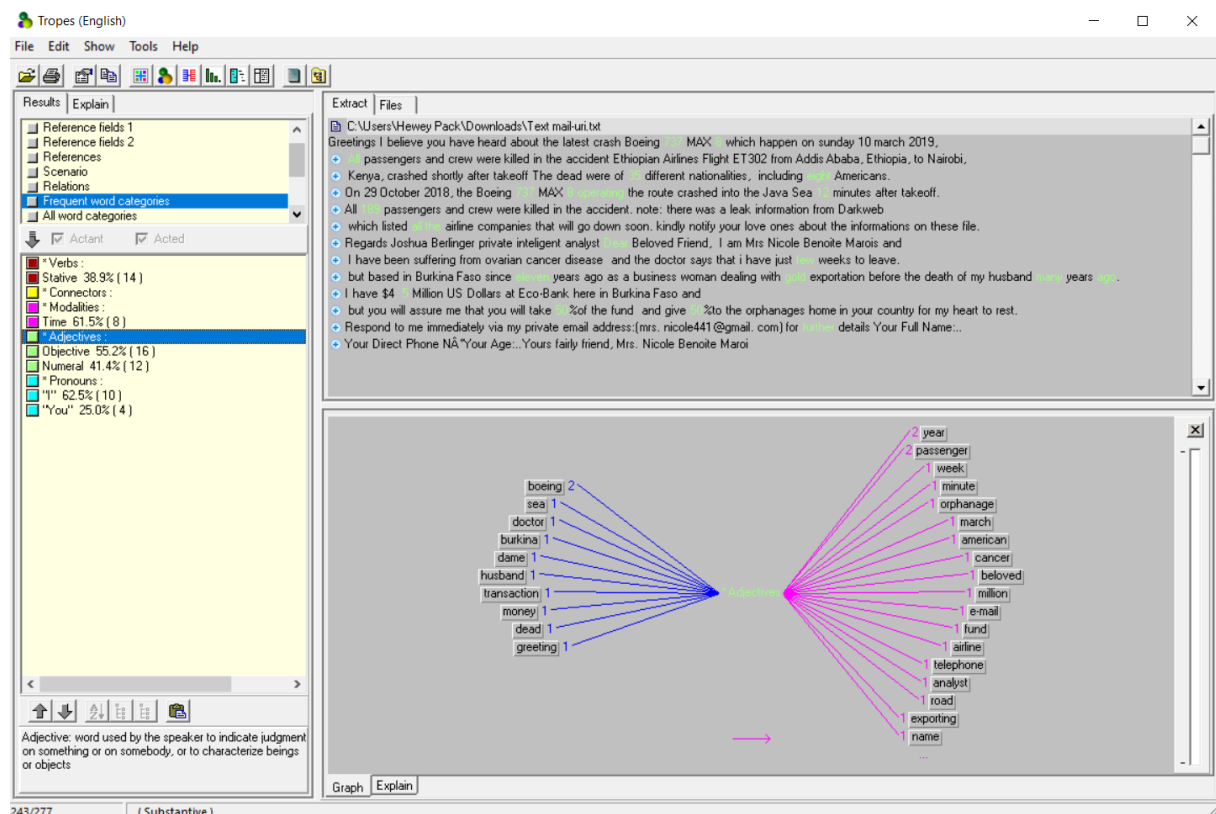


Fig.4. The relationship of adjectives in the phishing e-mail

Regarding the specialized representation of the categories of lexical units, the occurrence of certain nouns and adjectives is noted, which give a better characterization and knowledge of the reference context.

From the category of adjectives, the most used were used, sequentially, both objective and numerical, to increase the impact and determine an action, following the appeal to emotions (Fig.4).

Another important aspect is constituted by the presence of the pronoun of the first person, singular "I", which implies the idea of cohesion, inclusion, versus that of detachment, by using the third person, plural (Fig. 5). Many studies have shown that the use of the first person, singularly, in social interactions, implies differences in relative status or increased social connectivity.[8][9][10][11][12][13] In this case, compassion/ sympathy is called for, so that the action of the recipient takes place after the determination of the emotion.

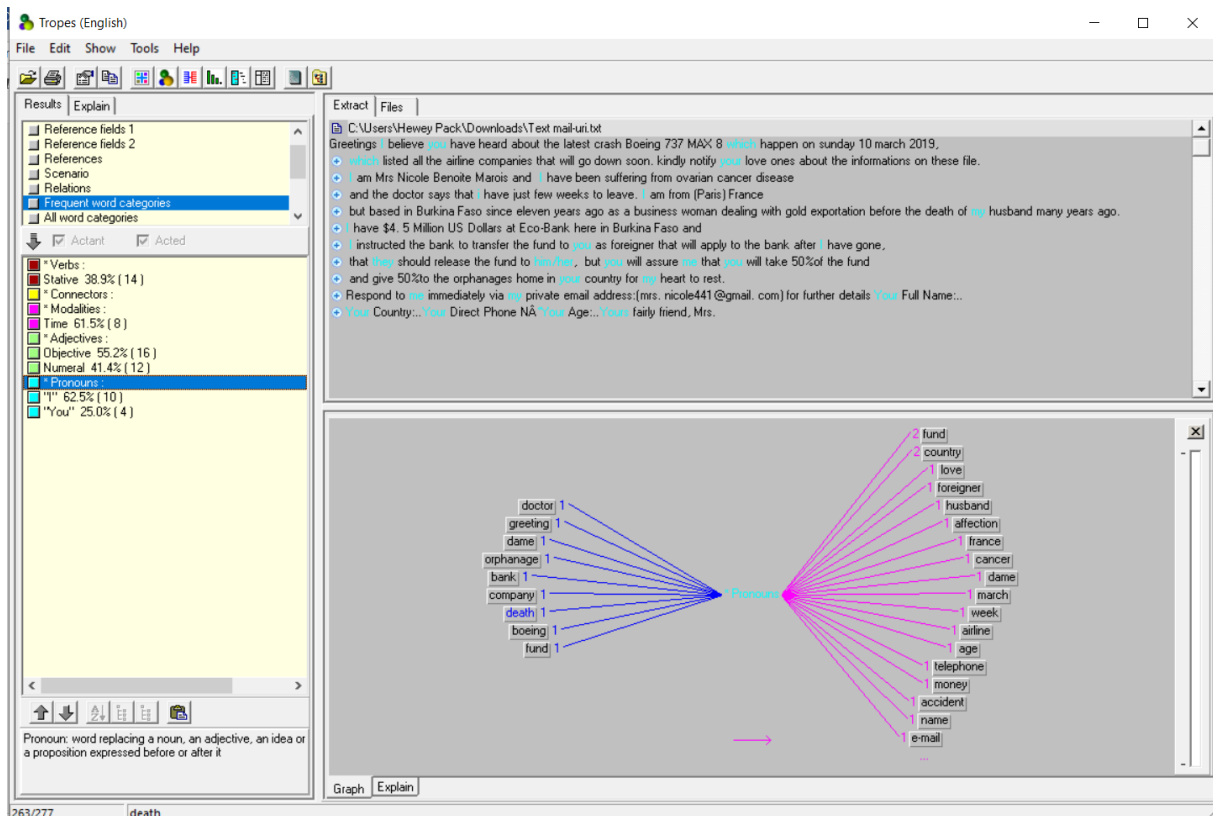


Fig.5. The relationship of pronouns in the phishing e-mail

#### 4. Conclusions

To summarize, it is important to emphasize that, despite the existence of complex software, an analysis to classify the method of social engineering used in the attack can only be done by a highly specialized person who has thorough specialist knowledge in quite varied fields. Interpretations provided by the software may not be used as such, in their raw form. They require specialized interpretation and integration by an expert in the field.

The Tropes app does have some downsides though. It cannot perform the comparative analysis of several texts (two sessions of the application must be opened simultaneously), and the semantic analysis in the case of small texts is not so relevant. Also, the semantic analysis in can be done in only 5 foreign languages: English, French, Spanish, Portuguese and Romanian.

#### References:

- [1] Chongyang BAI, Maksim BOLONKIN, Judee BURGOON, Chao CHEN, Norah DUNBAR, Bharat SINGH, V.S. SUBRAHMANYAN, Zhe WU, *Automatic Long-Term Deception Detection in Group Interaction Videos*, Published in 2019 IEEE International Conference on Multimedia and Expo (ICME), arXiv:1905.08617, <https://arxiv.org/pdf/1905.08617.pdf>, accessed on March 12, 2023
- [2] Paul EKMAN, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, Publisher W. W. Norton & Company, 1992, ISBN 0-393-30872-3
- [3] Carroll Ellis IZARD, *The face of emotion*, Publisher Appleton-Century-Crofts, 1971, ISSN 0271-9223
- [4] Ruxandra VRÂNCEANU, Constantin VERTAN, Răzvan CONDOROVICI, Laura FLOREA, Corneliu FLOREA, *A fast method for detecting eye accessing cues used in Neuro-Linguistic Programming*, Published in 2011 IEEE 7th International Conference on Intelligent Computer

- Communication and Processing, Publisher IEEE, 2011, DOI: 10.1109/ICCP.2011.6047873, <https://ieeexplore.ieee.org/document/6047873>, accessed on March 12, 2023
- [5] Richard BANDLER, John GRINDER, *The Structure of Magic – A book about Language and Therapy*, Publisher Science and Behavior Books, 1975, ISBN 08314-0044-7
- [6] Thomas R. COLE, *Using Technology to Separate Truth From Lies: A Review of the History, Science, Methodology, and Legal Implications of the Computer Voice Stress Analyzer*, Publisher University of Arkansas - Criminal Justice Institute, 2020, [https://www.cji.edu/wp-content/uploads/2020/07/Using\\_Technology\\_to\\_Separate\\_Truth\\_from\\_Lies.pdf](https://www.cji.edu/wp-content/uploads/2020/07/Using_Technology_to_Separate_Truth_from_Lies.pdf), accessed on March 12, 2023
- [7] Dan CARAGEA, Adrian CURAJ, *Analiza automată a discursului*, Publisher Publishing House of the Romanian Academy, 2013, ISBN 978-973-2299-2
- [8] Judee K. BURGOON, Tiantian QIN, *The Dynamic Nature of Deceptive Verbal Communication*, Published in Journal of Language and Social Psychology, Volume 25, Issue 1, March 2006, DOI: 10.1177/0261927X05284482, [https://www.researchgate.net/profile/Judee-Burgoon/publication/255542944\\_The\\_Dynamic\\_Nature\\_of\\_Deceptive\\_Verbal\\_Communication/links/53d659f40cf228d363ea4fd7/The-Dynamic-Nature-of-Deceptive-Verbal-Communication.pdf](https://www.researchgate.net/profile/Judee-Burgoon/publication/255542944_The_Dynamic_Nature_of_Deceptive_Verbal_Communication/links/53d659f40cf228d363ea4fd7/The-Dynamic-Nature-of-Deceptive-Verbal-Communication.pdf), accessed on March 17, 2023
- [9] Mark L. KNAPP, Roderick P. HART, Harry S. DENNIS, *An Exploration of Deception as a Communication Construct*, Published in Human Communication Research, Volume 1, Issue 1, September 1974, DOI: 10.1111/j.1468-2958.1974.tb00250.x, <https://files.eric.ed.gov/fulltext/ED081059.pdf>, accessed on March 17, 2023
- [10] Lina ZHOU, Judee K. BURGOON, Dongsong ZHANG, Jay F. NUNAMAKER, *Language dominance in interpersonal deception in computer-mediated communication*, Published in Computers in Human Behavior, Volume 20, Issue 3, May 2004, DOI: 10.1016/S0747-5632(03)00051-7, <https://www.sciencedirect.com/science/article/abs/pii/S0747563203000517>, accessed on March 17, 2023
- [11] Earl F. DULANEY Jr., *Changes in Language Behavior as a Function of Veracity*, Published in Human Communication Research, Volume 9, Issue 1, September 1982, DOI: 10.1111/j.1468-2958.1982.tb00684.x, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.1982.tb00684.x>, accessed on March 17, 2023
- [12] Kevin COLWELL, Cheryl K. HISCOCK, Amina MEMON, *Interviewing Techniques and the Assessment of Statement Credibility*, Published in Applied Cognitive Psychology, Volume 16, Issue 3, April 2002, DOI: 10.1002/acp.788, [https://www.academia.edu/23442499/Interviewing\\_and\\_Credibility\\_applied\\_cog\\_2002](https://www.academia.edu/23442499/Interviewing_and_Credibility_applied_cog_2002), accessed on March 17, 2023
- [13] Günter KÖHNKEN, *Statement Validity Analysis and the 'detection of the truth'*, Published in The Detection of Deception in Forensic Contexts, Publisher Cambridge University Press, 2004, DOI: 10.1017/CBO9780511490071.003, [https://www.researchgate.net/publication/280781159\\_Statement\\_Validity\\_Analysis\\_and\\_the\\_'detection\\_of\\_the\\_truth'](https://www.researchgate.net/publication/280781159_Statement_Validity_Analysis_and_the_'detection_of_the_truth'), accessed on March 17, 2023