A Prototype Theory-Based Study of Crimes in the English and Arabic Societies Using Web-as-Corpus

Fadia A. A. M. Badawi¹, KhaledElghamry², FayrouzFouad³

English Language Department, Faculty of Al-Alson, Ain Shams University, Cairo, Egypt

¹fadia_ahmed@rocketmail.com
²elghamryk@gmail.com
³fayrouzff@yahoo.com

Abstract: Drawing on the classified prototypes of semantic categories, this paper uses the principles of the prototype theory to cross-culturally explore the hierarchical prototypes of crimes in web-booted Arabic and English corpora. The study compares the conceptualization and categorization of crimes in the Arabic and English worlds. For doing so, four domains ‘com, org, info and edu’ are explored over the past seven years for differentiating the individual mentality from the institutional mentalities. The study uses the web as corpus to investigate the inductive pattern ‘crime* such as’ in three web domains: ‘.com’, ‘.org’ and ‘.edu’ in Arabic and English. The collected data is analyzed using AntConc software program. Indicated statistical tests are calculated to measure the universality of conceptualizing emotions across Arabic and English speaking worlds. The study explores the validity of using Rosch’s prototype theory (1975) in determining the dynamically changing categorization of criminal acts. It also tests the authenticity of using Web-as-Corpus, which is a very straightforward tool, in attaining so. Results reveal that there is an Arabic-English agreement on defining the basic levels of the concept crime. However, few sporadic differences exist and are subject of cultural differences between the oriental and occidental moralities. The principles of prototype theory are also valid as regards the metamorphosis of conceptualizing a concept at the folkloric and expert levels. This holds true as regards the Arabic and English data of this study.

Key words: Prototype Theory, Categorization, Corpora, Semantic Hierarchy, Cognitive semantics; Crimes.

1 INTRODUCTION

Categorization is central to our understanding as human beings and how we function in our daily lives. The process of categorization happens automatically and unconsciously. People are only aware of categorization if there is a problem. People tend to categorize every object or animal or even human beings around them; that is why the human conceptual system is so rich and complex. Categorization is related to cognitive sociolinguistics as the identity of a social group is determined by the changes and social roles of this specific group. The same goes for showing the similarities or differences between different groups or communities [1].

Alternative to the classical view on defining concepts, which dates back to Aristotle, that depicts concepts by necessary and sufficient features, prototype theory, formulated in the 1970s by Eleanor Rosch [2-4], assigns two basic principles that guide the formation of categories in the human mind. This dichotomy is based on the principle of cognitive economy and the principle of perceived world structure. Both pillars, where features differ in their relatedness to the concept, give rise to the human categorization system that has two dimensions: a horizontal and a vertical dimension. There, a category is a set of attributes that share characteristics of groups of people or objects, or ‘a number of objects that are considered equivalent’ [3, 5].

Notwithstanding, the category memberships created by humans are affected by demographical variables which are often culture-bound [6]. Rosch [5], however, contends although there are inconsistencies concerning category boundaries, people generally agree on good examples of a specific class. This agreement provides, according to Rosch, a universal categorization of concepts. The prototype is thus defined as the most representative example that comes to the mind when a person thinks of a category. The family resemblance constitutes the basis on which a few members are grouped together in a category. The study investigates how the prototype theory applies to the semantic categories and how the family members are extracted and ranked according to their basic and superordinate levels. The semantic category ‘crimes’ is explored in Arabic and English to highlight the differences and similarities between the two languages in the light of the prototype theory and its tenets [7].

The present study aims at using the Prototype Theory (PT), simply because it enables a mathematical grading of family members within a given category, to detect the Arabic and English prototypes of crimes. It also aims at evaluating the universality of conceptualizing crime at the folkloric (com-based) and expert (org-, edu- and info-) levels. Eventually, a bilingual hierarchical lexicon of crime is suggested.
2 PROTOTYPE THEORY AND SEMANTIC HIERARCHY

As one of the similarity-based theories of categorization, PT replaces the logical diachronic knowledge by a fuzzy statistically-based cognitive repository. According to Rosch [3], for a member to join a category, it does not have to satisfy a set of conditions, as suggested by the classical theory. It just needs to share some sort of family resemblance with the other family members (FMs) of the same category. Of these FMs, prototype is the most representative member of a category. The more a member resembles the prototype, the more representative of the category it is.

The PT assigns two basic principles that guide the formation of categories in the human mind. This dichotomy is based on the principle of cognitive economy and the principle of perceived world structure. Both pillars, where features differ in their relatedness to the concept, give rise to the human categorization system that has two dimensions: a horizontal and a vertical dimension [1-4]. Rosch [4] views concepts as containers of instantiations, together with the implicit assumption that conceptual combinations follow the set theoretical algebraic rules of classical logic.

The vertical dimension (basic-level categories), that is optimal for human beings in terms of providing optimum cognitive economy, relates to the mid-level detailing of inclusiveness of a particular category: the higher up the vertical axis a particular category is, the more inclusive it is. Inclusiveness relates to what is subsumed within a particular category. As such, the category CRIME is more inclusive than the category ASSAULT because it includes entities like FROGGERY, and MURDER in addition to ASSAULT. In turn, ASSAULT is more inclusive than SEXUAL ASSAULT (rape) because it includes other types of assaults in addition to rape (mugging, aggravated assault … etc.).

Categories higher up the vertical axis, which provide less detail, are called superordinate categories (hypernym). Those lower down the vertical axis, which provide more detail, are called subordinate categories (hyponym). It is assumed that humans categorize items by matching them against the prototype, or ideal exemplar, which contains the most representative features inside the category. Similarly, a prototype refers to a typical entity of its class or group. Categories are divided into superordinate, basic, and hyponymic levels [1-4]. Generically as it seems, individuals tend to accommodate the tokens belonging to a specific class in terms of distinctive features which form the set of information characterizing their typicality simply because they occupy the same semantic similarity space [2-5, 8].

Four features of prototypicality are proposed. First, prototypical categories exhibit degrees of typicality. Second, prototypical categories exhibit a family resemblance structure where semantic structure takes the form of a radial set of clustered and overlapping readings. Third, prototypical categories are blurred at the edges. Fourth, prototypical categories cannot be defined by means of a single set of criteria (necessary and sufficient) attributes [8]. Geeraerts [8] supports that corpus-based studies, which rely on frequency on its own, are working tools as regards negotiating conceptual structure. On the contrary, Taylor [9] considers frequency an unreliable determinant of prototypicality. For example, the typical color for children is sea blue even though they live far away from the sea and have not experienced the environment before. Consequently, Taylor [9] believes that children rely on their imagination to choose the best example of an instance. More recently, it is advised that extracting frequency-based category prototypes must be integrated with prominence, if indexing prototype effects is aimed, to be a precise operationalization of conceptual structure because interpreting results, which are based on frequency on its own as a unique index of conceptual-lexical structure, must be reviewed with great caution [10].

3 RELATED WORKS

Pramanik, et al. [11] study the pattern of crime using their proposed framework of automatic crime detection. The empirical evaluation of their model demonstrated adequate efficiency for criminal network discovery. Tabbert [12] used a corpus compiled from the British press to study the linguistic constructions used by criminals. He concludes that the criminals use linguistic constructions that specifically draw a contrast between them and the victims in a way that shows a dichotomous picture between the innocent victim and the evil criminal.

Drawing on the Prototype theory, Biria and Bahadoran-Baghbaderani [13] conduct a cross-cultural analysis of prototypicality norms used by male and female Persian and American speakers. Their results fail to support the idea of universality with regard to the conceptual organizations. However, they approve Rosch’s perspective on the fidelity of folkloric evaluation of central exemplars. Peripheral to these, disagreement relies on diversity in culture, original language and individual cognition.

Glynn[10] recruits the corpus methodology towards polysemy quantification for prototype theory. Frequency-based operationalization of prototypicality is considered profitable as regards distinguishing prototype effects in structure. However, properly applying fuzzy set theory to the analysis and properly interpreting these results in terms of prototype set theory requires further research.

Elshout et al. [14] conceptualize humiliation using their prototype-theory-based model. They compiled a user-generated content for characterizing prototypical humiliation involved feeling powerless, small, and inferior in critical situations. Putting the compiled criteria of defining humiliation, groups are tested accordingly. Results are
compared to expert-level verdicts. The authors provide compiling evidence on the significant similarity between folkloric and expert measures concerning prototypes of humiliation.

4 METHODOLOGY
This study uses AntConc content analysis program to quantitatively analyze two corpora representing Arabic and English categorization of crimes. AntConc is adjusted to extract a 5-word concordance to the left and right of the target pattern, lexicalizing the semantic category, in the Arabic and English corpora, respectively. That is to say, only collocates after the pattern ‘X such as’ are extracted in English. Thus, the study compares the most and least frequent collocates for each semantic category between Arabic and English.

A. Collection of Data
The data is a web-based collection of snippets from the web search engine GOOGLE. Some programs are used like EditPad Pro and AntConc to analyze the data. To collect the data, the pattern ‘X such as’ is used to collect snippets from GOOGLE and inserting them into EditPad Pro to compartmentalize the data. The domains that are analyzed are (.org, .com, .edu, .info). Plus, the study depends on analyzing 100 snippets per domain.

B. Description and Processing of Data
The analyzed data consists of two Arabic and English corpora representing one semantic category ‘crimes’. The Arabic corpus of crimes is 83,742 tokens. The English corpus of crimes is 812,286 tokens. Arabic and English data saved in UTF-8 files to facilitate digital processing.

EditPad pro and AntConc are used to analyze the data. EditPad pro is used to collect the data from Google and divide them into separate documents. AntConc is employed to show the concordance of every keyword to highlight the context in which the word is mentioned to facilitate the process of analysis. Also, the collocates of each word are highlighted through the program to emphasize the words that accompany the crimes in both cultures.

5 RESULTS
A. The English Prototypes of Crime
‘Theft’ is the most widespread crime in the com-based mentality which means this crime is rampant among individuals in the society. The same crime is not that rampant in the institutional domains such as ‘edu and info’. The same goes for assault which is the second most widespread crime. It is greatest in the com corpora but less in the institutional mentalities. On the contrary, when it comes to ‘rape, murder and robbery’, they score the least. The crimes exist in the society but a lot of research papers and studies tackle them. Crimes such as ‘terrorism, mugging, racketeering’ are only apparent in the organizational mentality (figure 1).

![Figure1. Web-based topper bootstrapped prototypes of crime in the four studied domains](image)

The most widespread crime is ‘theft’. Related to theft are robbery, violence, mugging and burglary. Assault is also at the core and it is surrounded by threats, violence, rape and robbery. Crimes that are related to body harm such as assassination, homicide, and violence are all grouped together under murder. In the same vein, all the
crimes that are related to money and economy such as 'racketeering, trafficking, and mugging fall under the umbrella of fraud. Thus, it seems clear that the family resemblance among the family members of the category crimes is conspicuous and they all go back to the prototype ‘murder’ (figure 2).

Figure 2. Web-based topper prototypes of crime latticed according to semantic relations. The dash line represents co-hyponymic relations among the bootstrapped FMs. The line length represents the typicality between the linked FMs.

B. The Arabic Prototypes of Crime
According to the Arabic data, ‘Murder’ is highest among individuals and organizations. This highlights the increasing percentage of violence among people in the Arab world. The same goes for ‘theft’ for both the society and organizations. However, they are mentioned slightly in both the info and edu mentalities. ‘Adultery’ is the third most widespread crime in the two domains and it is not even mentioned in the edu- or info-based mentalities. ‘Murder and violence’ are greatest in the edu-based mentality (figure 3).

Figure 3. Web-based topper bootstrapped prototypes of crime in the four studied domains.
Crimes, in the Arabic conceptualization, are related to each other through family resemblance. The prototype of the category is ‘murder’ and it is surrounded by all the family members based on their resemblance to the prototype. ‘Assassination, annihilation, assassination, assault, and theft’ are related directly to murder and violence. ‘Rape’ which is one of the most representative crimes is also surrounded by similar crimes such as ‘terrorist, torture, and harassment’. Moreover, ‘rape’ is at the core of the crimes and around it are crimes such as ‘kidnapping, assault, robbery’. Regarding ‘bribery’ which is one of the most widespread crimes in the Arab world, it is surrounded by crimes that are related to money such as ‘embezzlement, fraud, smuggling, and cheating’.

Figure 4. Web-based topper prototypes of crime latticed according to semantic relations. The dash line represents co-hyponymic relations among the bootstrapped family members.

C. Contrasting Prototypicality of Crime in English and Arabic

After sorting all the web-based crimes, the hypernym and basic-level domain of every FM is determined using Roget’s classification. Thus, stealing, killing, evil, falsehood, impurity, attack and destruction are the topper basic-level domains. To such basic terms, several co-hyponyms are affiliated. This holds true in Arabic and English. Hierarchically implicating, both cultures reflect, through linguistic parameters, similar cognitive profile of conceptualizing crime – semantic hierarchical cluster (Table 1).

To measure the frequency of each domain, the following equation is used:

\[ Intensity = \sum_{i=1}^{n} x_1, x_2, x_3, \ldots, x_n \]

where \( x \) denotes the co-hyponyms in every basic-level domain.
TABLE 1
ORGANIZATION OF COGNITIVE SEMANTIC HIERARCHY OF CRIMES IN ARABIC AND ENGLISH

<table>
<thead>
<tr>
<th>Co-hyponyms</th>
<th>Basic</th>
<th>Co-hyponyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>اختطاف، إنزاز، سطو، اختلاس، خطف، سرقة</td>
<td>Stealing</td>
<td>Carjacking, Hacking, Abduction, Blackmail, Burglary, Embezzlement, Extortion, Kidnapping, Larceny, Looting, Poaching, Pickpocketing, Piracy, Robbery, Shoplifting, Stealing, Theft</td>
</tr>
<tr>
<td>اختطاف، ابنف، سطو، سرقة</td>
<td>Killing</td>
<td>Assassination, Genocide, Homicide, , Infanticide, Manslaughter, Murder, Shooting, Stoning, Suicide, Thuggery</td>
</tr>
<tr>
<td>ارتهاب</td>
<td>Evil</td>
<td>Abuse, Damage, Mischief, Terrorism</td>
</tr>
<tr>
<td>اغتيال</td>
<td>Fighting</td>
<td>Assassination, Genocide, Homicide, , Infanticide, Manslaughter, Murder, Shooting, Stoning, Suicide, Thuggery</td>
</tr>
<tr>
<td>اغتصاب</td>
<td>Impurity</td>
<td>Adultery, Prostitution, Rape</td>
</tr>
<tr>
<td>اعتداء</td>
<td>Attack</td>
<td>Aggression, Assault</td>
</tr>
<tr>
<td>تخراب</td>
<td>Destruction</td>
<td>Destruction, Vandalism</td>
</tr>
</tbody>
</table>

The prototypes of crime in Arabic and English are therefore defined in Table 2. Although the ranking is slightly different, the p-values which correspond to the calculated t-test and Spearman correlation coefficient are greater than 0.05. The null hypothesis is thus accepted. This indicates that there is no statistical significance for the bilingual difference in populating and ranking the prototypes of crime.

TABLE 2
ARABIC AND ENGLISH TOPPER FAMILY MEMBERS OF CRIMES

<table>
<thead>
<tr>
<th>Rank</th>
<th>English Prototype</th>
<th>Arabic Prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Theft</td>
<td>قتل</td>
</tr>
<tr>
<td>2nd</td>
<td>Fraud</td>
<td>سرقة</td>
</tr>
<tr>
<td>3rd</td>
<td>Assault</td>
<td>زنا</td>
</tr>
<tr>
<td>4th</td>
<td>Murder</td>
<td>فساد</td>
</tr>
<tr>
<td>5th</td>
<td>Rape</td>
<td>اغتصاب</td>
</tr>
<tr>
<td>6th</td>
<td>Robbery</td>
<td>رشوة</td>
</tr>
<tr>
<td>7th</td>
<td>Burglary</td>
<td>تخراب</td>
</tr>
<tr>
<td>8th</td>
<td>Violence</td>
<td>ارهاب</td>
</tr>
<tr>
<td>9th</td>
<td>Trafficking</td>
<td>تزوير</td>
</tr>
<tr>
<td>10th</td>
<td>Abuse</td>
<td>عرف</td>
</tr>
<tr>
<td>11th</td>
<td>Embezzlement</td>
<td>تحرش</td>
</tr>
<tr>
<td>12th</td>
<td>Mugging</td>
<td>سطو</td>
</tr>
<tr>
<td>13th</td>
<td>Terrorism</td>
<td>اختلاس</td>
</tr>
<tr>
<td>14th</td>
<td>Racketeering</td>
<td>احتيل</td>
</tr>
</tbody>
</table>

D. Contrasting Crime Prototype across the Four Studied Web Domains

We provide representative concordancing of the prototypical crime, theft, from our compiled corpora to reflect the different perception of the same prototype across the studied communities.

COM
describe computer crimes, such as theft of hardware, software, intellectual property, time and services,
Similar to crimes against wildlife, agricultural crimes such as theft of agrichemicals, cattle and horse rustling,
Predatory crimes, such as theft, during which the criminal does not directly confront the victim.
Petty crime definition are minor crimes such as theft, trespassing etc
feel particularly vulnerable to crimes such as theft and assault. A common form of violence against older people
Crimes such as theft and murder were very common during the medieval ages. Corporations often do not officially report other crimes such as theft and hacking. Inflation in price is believed to drive addicts to commit crimes such as theft and robbery, which are damaging to the economy. Property crimes such as theft, robbery or sexual assault are sometimes very prevalent at any school. How Iran deals with crimes such as theft, fraud, forgery, insult, crimes against national security, crimes against the state, and even the advance in technology which facilitates committing the crimes. Correlation lines show the differences between the four mentalities when it comes to the perception of crimes. The com mentality is related to common people (folkloric level). In this mentality, the emphasis is on the crimes that are committed by people and affect the people on the street. That is why crimes such as theft, fraud or assault are dominant. Also, the concordance lines revolve around explaining who people blame. People blame governments, the economic conditions and even the advance in technology which facilitates committing the crimes. However, in the org mentality, which is the organization perception of crimes such as theft, the focus is on the factors behind the crimes and differentiating between the different types of theft. For example, identity theft is the crime that most organizations are interested in as shown in the concordance lines.

The above concordance lines show the differences between the four mentalities when it comes to the perception of crimes. The com mentality is related to common people (folkloric level). In this mentality, the emphasis is on the crimes that are committed by people and affect the people on the street. That is why crimes such as theft, fraud or assault are dominant. Also, the concordance lines revolve around explaining who people blame. People blame governments, the economic conditions and even the advance in technology which facilitates committing the crimes. However, in the org mentality, which is the organization perception of crimes such as theft, the focus is on the factors behind the crimes and differentiating between the different types of theft. For example, identity theft is the crime that most organizations are interested in as shown in the concordance lines.

Regarding the info domain which is concerned mainly with statistics about crimes, the concordance lines clearly show the numbers and percentages of the occurrence of theft and other prevalent crimes. Also, this domain is interested in the doers of the crimes and the people affected by the crimes (criminals and victims). Thus, there is mention of the possibility that people who commit crimes now like theft or fraud, will likely commit more dangerous crimes in the future. Furthermore, the edu mentality, which is the educational perception of crimes, is mainly concerned with research papers and the studies that are done about the crimes. Also, it is concerned with the place and time of the crimes to explain why such crimes happen in certain environments. Thus, the concordance lines prove that there are truly differences among the different mentalities in the same culture not just across cultures as has been show earlier.

6 DISCUSSION

Recruiting the prototype theory to cross-culturally explore the hierarchical prototypes of crimes in web-booted Arabic and English corpora, this study compares the conceptualization and categorization of crimes in the Arabic and English worlds. Results show that both Arabic and English prototypes hinge on the same basic levels: stealing, killing, evil, falsehood, impurity, attack and destruction.

In spite of the similar conceptualization of crime, there are some individual differences. For example, the folkloric (com-based) perception of crime prototypes, in English and Arabic differs from that of the expert (org-,
edu- and info-) peer. ‘Theft’ is the most widespread crime in the English com-based mentality which means this crime is rampant among individuals in the English society. The same crime is not that rampant in the institutional domains such as ‘edu and info’. The institutional domains are not that affected by this crime as the people are. The same goes for assault which the second most widespread crime. It is highest in the com corpora but less in the institutional mentalities. On the contrary, when it comes to ‘rape, murder and robbery’, it seems that the institutional mentalities are focused on these crimes more than the individuals. The crimes exist in the society but a lot of research papers and studies tackle them. Crimes such as ‘terrorism, mugging, and racketeering’ are only apparent in the organizational mentality.

‘Murder’ is highest among individuals and organizations. This highlights the increasing percentage of violence among people in the Arab world. The same goes for ‘theft’ for both the society and organizations. However, they are mentioned slightly in both the info and edu mentalities. ‘Adultery’ is the third most widespread crime in the two domains and it is not even mentioned in the edu- or info-based mentalities. ‘Murder and violence’ are greatest in the edu-based mentality which highlights the interest of the researches in these two crimes as they are also widespread among people. Characteristic to the Arabic culture is ranking ‘زنا/ Adultery’ in the third most widespread crime. However, it is not prototypically criminalized in the English society. This highlights a sporadic cultural difference among the Arabic and English societies.

7 CONCLUSIONS AND FUTURE WORK

According to the findings of this study, there is an Arabic-English agreement on defining the basic levels of the concept crime. However, there are some differences between the two cultures when it comes to the most widespread crime in Arabic and the most widespread crime in English. The two principles of the prototype theory proved effective when they have been applied to the data. Also, the differences between the four communities of practice (com, info, edu and org) are accentuated to highlight the differences in the same culture and across cultures. The cross-cultural difference between Arabic and English categorization of crime prototypes seems to be of a theological background.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their helpful comments on an earlier draft of this manuscript.

BIOGRAPHY

Fadia Ahmed A. M. Badawi is now an Assistant Lecturer, Faculty of Al-Alsun, Ain Shams University, Egypt. Badawi’s research interests include cognitive semantics, corpus studies, literary linguistics and ethnopragmatics. Badawi obtained her MA in Linguistics in 2017.

Dr. Khaled Elghamry is an Associate Professor of (Computational) Linguistics, Faculty of Al-Alson, Ain Shams University, Egypt. Dr. Elghamry obtained his Ph.D. in Computational Linguistics, 2004, Indiana University, USA. He was a visiting scholar, Department of Linguistics, University of Florida 2007-2010. Dr. Elghamry is the Co-Founder of the Midwest Computational Linguistics Forum, Indiana University, and also the Co-Founder of the Arabic Digital Content Statistical Database, and the Arabic Digital Content Foundation Report. Dr Elghamry presented and published research in international conferences and journals on different technical issues in the automated processing of the Arabic language and content, sentiment analysis, and tracking of online public opinion.

Dr. Fayrouz Fouad is a Lecturer in Linguistics, Faculty of Al-Alson, Ain Shams University, Egypt. Dr. Fouad obtained her Ph.D. in Linguistics in 2008. Her research interests include theoretical linguistics and critical discourse analysis.
REFERENCES

TRANSLATED ABSTRACT

دراسة ذخائرية للتصنيف التراثي للجرائم في اللغتين الإنجليزية والعربية باستخدام نظرية النموذج الامثل

فادية أحمد بدوى – أم خالد الغمري – د. فريروز فؤاد

قسم اللغة الإنجليزية – كلية الإنسانية – جامعة عين شمس

Fadia_ahmed@rocketmail.com
elghamryk@gmail.com
fayrouzff@yahoo.com

ملخص:

يهدف البحث إلى الربط المعاني بين اللغة والمجتمع من خلال استنباط كيفية تأثير اللغة على التفكير المجتمعي. و يقوم البحث على تحليل مفاهيم جرائم شائعة في اللغتين الإنجليزية والعربية، حيث تم تعلم المفردات المتشابهة في اللغتين من حيث ترتيب أشكال الكلمات المتشابهة لهذا المفهوم في اللغتين العربية والإنجليزية. ويستفيد البحث من نتائج الدراسة التي أظهرت أن الجرائم الأكثر شيوعًا في المجتمع العربي هي القتل بينما الجريمة الأكثر شيوعًا في المجتمع الغربي هي السرقة. وظهرت النتائج أيضًا أن الجرائم المجتمعية في العرب و угمة الجريمة بال gameIdj ينتمون إلى العالم العربي وتتعلق بالعنف والإيذاء الجسدي.

الكلمات الدلالية: النموذج الامثل-جرائم مجتمعية-دراسة ذخائرية-تصنيف