International Journal of Social Science Archives

ISSN: 2707-8892

Available at *www.ijssa.com*

International Journal of Social Science Archives, July-September, 2024, 7(3), 275-300

Extraction of Elementary Meaning Units (EMUs): Understanding Culture through Semantic Association

Mahpara Ayaz[a], Dr. Mujahid Shah[b], Sawaira Maula Dad[c]

[a]M.Phil. Scholar, English (Linguistics), Abdul Wali Khan University Mardan. [b]Associate Professor, Abdul Wali Khan university. [c]M.Phil. Scholar, English (Linguistics), City University of Science and Information Technology, Peshawar

[*]**Email:** mujahidshah@awkum.edu.pk

***Abstract*:** Culture is a complex and variable social and semiotic construct comprising shared explicit and implicit behavioral, intellectual, and ethical patterns learned through synchronic and diachronic processes of transmission and socialization. The main objective of this research initiative is to increase our understanding of the cultural structures and the relationships among written content, semantics, and culture. A corpus methodological approach was used in this study to analyze culture through semantic association. The "projective methodology" was used to collect data using an indirect questionnaire. This research utilized corpus-based approach and relied on positivist/empiricist framework. Wmatrix software was used to analyze first question. The Molinaris'index metric was employed to determine the levels of conventionalization of the semantic domains of the corresponding node terms. Fleischer's theory of the cultural system was applied to analyze the data in order to comprehend culture. The various levels of semantic domain dominance in Pakistani culture was demonstrated by using elicited data extracted from people of Pakistan. This study will be of immense importance since it will use corpus software to extract elementary meaning units (EMUs), which will help people to understand culture. Its significance also stems from the use of the corpus from other angles, including marketing research.

Key words: Semantic association in culture, Wmatrix, Level of conventionalization, Elementary meaning Unit

1. Introduction

Culture is an intricate and variable, communal and semiotic construct comprising of shared overt and embedded behavioral, intellectual, and ethical patterns learned through synchronic and diachronic processes of transmission and socialization (e.g., nationality, common interests, common interactional patterns, and membership in a professional category etc.) While certain cultural qualities may be expressed through language or other produced goods, others may not. Though its members aren't usually conscious of this diversity, distinct cultures create and share different characteristics. People's positions, responsibilities, shared beliefs, knowledge, and values shape the reality we live in, is the culture (Halliday, 1978).

Raymond Williams was renowned for his analysis of lexical items with meaning in understanding reality. In "Culture and Society," he remarked that certain lexical items take on special meanings significant in different eras.

He believed that these expressions indicate alterations in how people see reality. He stressed the five sentences below in particular: "Culture, industry, art, class, and democracy are all intertwined. In his famous book "Keywords, In A Vocabulary of Culture and Society" (Williams 1976), he presents the idea that language and how people use it may be better understood by contrasting the meanings of certain relevant terms with those of people's everyday experiences, or "Keywords." He gives instances of concepts that he believes are related to the notion of culture (and society) and provides insight into its complicated meaning. Williams says that we analyze culture by employing several new notions connected to it. Thus, there is a connection between corpora, semantics, culture, and language.

1.1 Background of the Study: Culture, Language, and Semantics

Literature, art, archaeology, philosophy, anthropology, semiotics, and, more recently, linguistics, translation studies, and marketing have all contested how to define culture. Despite the differences in how each scientific field conceptualizes culture and the specificities of individual theories, several basic principles appear to be held by the scientific community at large. According to them, culture is a highly complex and diverse social and semiotic event that is acquired through information transmission mechanisms; it may evolve both diachronically and synchronously; and certain cultural features can be seen in everyday life through language or other man-made objects. However, two features of culture are particularly relevant here: it arises through language and is widely shared by individuals of the same group.

Indeed, language is our fundamental way of describing the environment and communicating our thoughts and ideas. Beliefs and judgments, as well as values and value orientations, are part of what Hall (1989) refers to as "the informal level of culture." This level of culture which is separate from the other two levels: the technical and the formal ones, can neither be taught nor studied; it is handed on and gained instinctively or out of awareness. People generally react in ordinary life and conversation at this casual level. Even if we are unaware of it, beliefs, values, and value orientations pervade our thoughts and regulate our mental and bodily activities. As a result, they pervade our language on all levels, from semantics to grammar, pragmatics to discourse structure. The degree of semantics was considered in this work. The notion of Elementary Meaning Units (or EMUs) was utilized as inspiration, a term initially developed by Greenberg (1966, quoted in Maday, B. C) and later employed in anthropology for cross-cultural comparisons. Elementary Meaning Units are personal meaning reactions to certain lexical items.

Computationally, it is practical and accurately represents languages being systematically processed by humans and machines**. CORPUS stands for (C) capable (O) f (R) representing (P) potentially (U) unlimited (S) election of texts** (Dash, 2005).

The most significant advantage of corpus-based linguistics is the value contributed by annotation at practically all levels of language, from syntactic annotation to semantic annotation and part of speech tagging. These annotations allow for a more in-depth examination of language and make data processing easier. Another advantage of CBA is the mix of intuition and statistics, which many researchers may see as an easy approach while conducting research. Furthermore, with a corpus-based method, the size of a corpus is more variable, as an exceptionally large size is not necessary. It is considered a balanced corpus as long as it is sampled according to particular rules to be appropriately representative. Considering the above instances of researchers about approaches to corpus linguistics, the current research takes into account the corpus-based approach.

Even though corpora and quantitative analytical methods can be adopted for this purpose, the low proportion of cultural studies in the corpus compared to qualitative cultural studies seems to indicate that culture has been a relatively understudied area in this particular branch of linguistics. Selecting a distinct cultural theoretical framework is essential for an accurate and convincing interpretation of the facts, just like in any other type of scientific investigation. This study uses corpus linguistics particularly corpus utilizing corpus analysis tools as its methodological framework and theories of culture, namely Fleischer's theory, as its theoretical foundation. The primary goal being pursued by the researcher is under examination.

1.2 Objectives

a) To highlight the semantic association of lexical items orange, tea, tradition, and alcohol by applying corpus analysis.
b) To find out levels of conventionalization for lexical items using corpus analysis.
c) To accentuate the dominant semantic association in understanding culture through Fleischer's concept?

1.3 Questions

a) What are the semantic associations of the lexical items "orange, tea, tradition, and alcohol?
b) What are the levels of conventionalization of lexical items "orange, tea, tradition, and alcohol?
c) How is semantic association linked with culture through Fleischer's concept?

1.4 Statement of the Problem

Very limited research in Pakistani context has been carried out in terms of corpus, semantic and culture. This research will endeavor to find semantic association through the extraction of EMUs in an understanding of a single culture only.

1.5 Significance of the Study

This research will be of immense significance to all humans for understanding culture through semantic association using corpus software. Further, the methodology used in understanding a culture will be very effective in supporting certain cultural theories. Apart from this, the methodology applied will be very beneficial in marketing research through the extraction of EMUs by analyzing various semantic associations.

1.6 Delimitations

Due to limited time, the researcher limited its study to four nodes that are: alcohol, tradition, orange and tea", semantic association particularly semantic domains, along with semantic tagging, levels of conventionalization and single cultural theory. Further, the study was limited to the Pakistani context only.

2. Literature Review

2.1 A subjective study of culture through extraction of Elementary Meaning Units (EMUs)

Subjective culture, also referred to as implicit culture, was initially studied by anthropologists Szalay and Maday (1973). They began by looking at the connections between languages and how language groups respond vocally. These authors define culture as a "group-specific cognitive organization or world picture" made up of psychologically meaningful components. The words that are triggered by specific other words have psychological meaning embedded in them (to extract elicited data through free word association) known as EMUs (Elementary Meaning Units). EMUs are the subjective meaning reaction to an individual word (Greenberg, 1960). Composition, organization, and dominance are important elements in cultural studies. Numerous factors contribute to the psychological meaning, including the context of use, imagery, emotional responses, functions, brand, marque, etc. To determine the two groups' respective cultural priorities, the participants were requested by the researchers to create a compilation of twenty-five important life categories (domains) and provide as many linked answers for each of them as possible. Two lists that were unique to each culture were created by combining the lists of the Korean and American learners. Among the subjects that were frequently utilized were words that serve as stimulus for a free verbal association task were education, manners, and family. Each culture's domains, both those that occur frequently and those that do not, were submitted to content analysis. During the content analysis phase, fewer categories were employed to classify responses. Researchers from both cultures manually detected and categorized EMUs. To compare domains across different groups and perceive each domain's cognitive structure, affinity indices were created between the EMUs of each participant group for each word used as a stimulus. The connections of affinity between domains (such as family, manner, and education) were assessed once the affinity structure and cognitive organization for each domain word were established. The writers noted that "words utilized in the portrayal of a specific topic in general reflect the same enduring cultural tendencies" for each set of participants. This indicates that the affinity between each group and the other domains is consistently either low or high. Regarding the familial and educational domains, politeness, civility, manners, and bowing (EMUs in the manner domain) continually demonstrated a good affinity for Korean students but a weak affinity for American students in the experiment. Based on these findings, the writers postulated that a limited number of carefully chosen EMUs would be adequate to uncover broadly applicable cultural patterns across a wide range of semantic domains.

2.2 Comprehension of Culture via Levels of Conventionalization

Wilson and Mudraya's (2006) paper supports Fleischer's findings. The two aforementioned linguists were inspired by Fleischer's theory, which holds that despite the fact that judgements involve both individual as well as cultural elements, a concept or object being judged can be thought of as having conventionalized as a shared symbol if

judgements from a variety of individuals tend to overlap. As a result, they decided to "consider one potential link across the onomasiological and cultural levels," using quantitative approaches. i.e., to investigate the existing interaction between the cultural and onomasiological levels. Their research was based on elicited verbal responses in Russian that were gathered through two distinct activities. The participants in the first test, who were all native Russian speakers, were shown 12 photos representing various types of shoes and asked to identify each one. The second assignment required the participants to complete 10 instances of the identical sentence like "I believe that the woman who is wearing these shoes". A modified version of a well-known self-identity study exercise was the second task, which looked for distinctive traits shared by individuals who wear a certain shoe style. An evenness index, which, like TTR, takes into account the relationship between types and tokens as well as the evenness of token distribution among types, was applied to count and analyze the names provided to shoe styles and qualities. In contrast to the authors' working hypothesis that "where participants agreed more easily regarding a suitable title [...] would likewise [...] elicit greater conventionalized connections about the possibility of wearers" in the Russian context, there was certainly no apparent connection between mentioning uniformity and the evenness of connections. While earlier research only obtained information within the selected conceptual framework, Wilson & Mudraya's study uses evidence to test theoretical hypotheses as well as to the quantitative techniques that were used. Furthermore, although linguists conducted both this and Fleisher's earlier research, their methodologies are more like to those of other fields of study.

Fleischer's study from 2002 sought to determine the semantic profiles of beverages in three distinct countries—France, Germany and Poland—and evaluate the degree to which their representations had become customary in those cultures. When he speaks about conventionalization, he means the degree to which every member of a given culture shares the semantic characteristics of a single type of beverage. To accomplish this, he asked a rather generic question around the lines of, "What immediately comes to mind whenever you look at each of the titles of beverages on the list?" and separated the participants into three distinct categories, one representing each of all three cultures. Fleischer's established a "corpus" of spontaneously generated vocal evaluations of drinks utilizing data collected by the participants. Semantically, the responses of the responder were examined. Three themes—characteristics and connotations "(Konnotationen/Images), trademarks and proper names (Umschreibungen/Marken)", and evaluations—were identified. Then, each theme's words and phrases were organized into unnamed semantic groups. Hapax legomena were disregarded since they were thought to be related to personal preferences and sentiments rather than to societal or cultural orientations after that, the categorical features referred to as Types and unique examples of every characteristic referred to as Tokens were used to determine the Type-Token Ratio (TTR) of the responses. Then, using confidence intervals based on averages and standard deviations, three levels of conventionalization— "high, medium, and low"—were created. As a result, TTR was seen as a measure of intracultural conventionalization of alcohol imagery. The findings demonstrated that different beverages had varying degrees of conventionalization in the three cultures taken into account. So, for instance, all three cultures exhibit a high level of conventionalization in the production of chocolate and milk. This indicates that, within each nation, a significant portion of the semantic connections of a particular beverage are shared by every individual in that culture, or, to put it another way, that pictures of milk and cocoa are mainly shared by all individuals within culture. However, beer, Coke, and coffee exhibit low, high, and medium levels of conventionalization, respectively, in Poland, Germany, and France. Correlation tests were used to illustrate intercultural disparities in conventionalization. Finally, yet importantly, each drink's semantic characteristics are explained and explored through historical, social, and more broadly cultural occurrences. The clear connection that Fleischer's study makes between words (drink names), linguistic connotations (drink descriptions), and cultural emblems, in addition to the effort to measure conventionalization, make it particularly fascinating (Type-Token Ratio)

3. Methodology

This chapter gives a brief overview of the methodology, the corpus, and the procedures used to create the corpus and sub-corpora using elicited data, as well as why it was important for understanding the culture and creating own corpora through elicited data. The present study incorporates both theoretical and methodological elements. The study is methodological since it uses corpus analysis techniques and has a theoretical foundation in a conceptual framework for cultural diversity.

3.1 Research Design

The research paradigm is defined as "a framework incorporating all the established beliefs about a subject, a structure of what direction research should be taken, and how it should be executed" (Shuttleworth, 2008). Positivists and empiricists think that mathematics and statistics may explain an objective world. This institution employs quantitative research techniques (William, 1976). According to Howell (2013), positivism is a product of the empiricist school of thought which maintains that hypotheses can only be proven true by experimental means. Ryan (2018) continues, "The study of objects is to identify these links and to justify them scientifically using logical instruments." He claims that positivism is consistent with the notion of natural science and that there must be logical relationships both inside and between objects. Positivism, which is frequently connected to experiments and quantitative research, is seen as an evolution or kind of empiricism. According to Phillips and Burbules (2000), empiricism is one of two philosophical schools under foundationalism—empiricist and rationalist —that holds that knowledge ought to be impartial and unaffected by the researcher's attitudes and opinions. Based on the above studies the research design is quantitative in nature.

3.2 Participants

The study is confined to Pakistani participants. Convenience sampling, a type of nonprobability sampling, was employed to incorporate participants in the research who meet certain practical needs, including being reachable, available during a specified period, or willing to take part (Fink, 2003). It also addresses research on a population that is accessible to researchers. The main premise of convenience sampling is that the members of the target population are all the same. In other words, the results of a study conducted using a random sample, a sample from the nearby region, a sample from a voluntary participant group, or a sample drawn from a distantly populated area would all be equivalent. The survey was conducted online. The questionnaire was sent by email, messenger, and WhatsApp to those who responded. WhatsApp, email, messenger and Facebook are the only practical methods for gathering data that are readily available, cost-effective, and time-efficient. The target group received links to the surveys that were designed using Google Forms. The target group age ranged from 20 years to 30 years irrespective of cast, religion, language and color. The questionnaire was send to groups like CSS/PMS, Federal Public Service Commission and World Times Institute. Equal number of respondents were taken from each group.

3.3 Data Collection

Sentence completion and sentence writing activities on questionnaires were used to precisely gather the elicited data for this investigation. Hair, Bush and Ortinau (2009) and Wilson and Mudraya (2006) inspired the questionnaires' formation. For current research, data was collected through an indirect questionnaire using a method called the “projective technique “which reflects the respondent’s true beliefs and opinions through “free word association” and “sentence completion”. Data on the four node words "orange, tea, tradition, and alcohol" were obtained through the completion of questionnaires that asked respondents to complete sentences and create phrases using related node keywords (Bianchi, 2010). Two questionnaires were prepared to collect the elicited data. The first questionnaire contains the node words **“ORANGE” and “TEA”.** The questionnaire was designed that include sentence completion and writing sentences for the free word association. The second questionnaire consists of two node words **“TRADITION” and “Alcohol”**. Following the research, participants were required to come up with an additional 10 phrases using appropriate free words.

Table 1: General information about elicited data and no. of participants

Corpus	**Respondents**	**Total words**
Overall corpus	158	35,023
Alcohol corpus	60 respondents	5,179

Tradition corpus	60 respondents	9,495
Orange corpus	99 respondents	9398
Tea corpus	99 respondents	10,885

Ninety-five percent of the whole population replied to the surveys. The first survey had responses from sixty people, while the second survey had responses from ninety-nine respondents. A corpus of 35,023 (thirty-five thousand and twenty-three) words was produced by gathering primary data, also known as elicited data. Four corpora were created from the corpus:
The "orange corpus," which has 9398 words; the "tea corpus," which has 10,885 words; the "tradition corpus," which has 9,495 words; and the "alcohol corpus," which has 5,179 words.

3.4 Corpus Formation through Elicited Data

The creation of the plain corpus text involved the following steps

1. All the data was extracted and converted to Word format.
2. Each corpus was run through an IPVOID to eliminate unnecessary characters like punctuation (https://www.ipvoid.com/remove-punctuation/).
3. Pashtu and Urdu terms from the elicited data that were used by respondents to complete the questionnaires were translated into English.
4. Many of the respondents' often-used short versions of terms were also modified in addition to this.
5. All the texts were then saved in plain text format.

3.5 Theoretical Framework for Culture Analysis

Fleischer's theoretical cultural notion was deemed suitable for the investigation of the question.
Fleischer simplified the idea and made the following claims:
“Both collective and discursive symbols are made up of three elements, which Fleischer calls core, current field, and connotation. The core is a stable element. Collective symbols with long-standing have a strong, dominant core. The current field is a rather generalized, but not yet stabilized element. Both core and current fields are expressions of cultural meanings. Finally, the connotation field is an expression of individual meaning and as such has no stabilization at all; it is connected to the particular lexical meaning’’(Fleischer, 2002).

4. Analysis

4.1 Extraction of Semantic Domains from Elicited Data

In this chapter, the elicited datasets are used to investigate three distinct approaches. At the semantic domain level, the concepts of alcohol, tradition, oranges, and tea are analyzed. The first method generates concordances for every word, carefully examines the concordance lines, and allocates each word to one or more of the accessible semantic categories. This allows automatic semantic analysis to be applied to the most frequently occurring content lexical items in the elicited data. Studying the concordance lines, each word is assigned to one or more of the available semantic categories. In the second, sentences are extracted from the automated dataset using the most frequently occurring content terms. Thirdly, stop-lists are utilized to automatically eliminate words that appeared among the most frequently occurring items but were not included in any of the semantic categories. Examples of these words include function words and other highly frequent words that were not desired. More precisely, each dataset was given its unique stop-list, which was built for it. Among the stop-lists that are used—which are modifications of stop-lists made for study—are modal verbs, conjunctions, relative pronouns, articles, prepositions, adverbs of place and time, auxiliary verbs (all forms of auxiliary verbs), and other several forms of the same word. Personal pronouns and adjectives such as her, him, her, and his are not excluded since they fall under the semantic categories

of MEN and WOMEN, nor are modal verbs. It turned out that semantic issues needed to be resolved by concordance analysis. After each semantic domain's concordances were created, matching was completed by going over each concordance line. Following the process for creating plain text corpora, the corpora were passed through Paul's (2008) Wmatrix 5 to extract semantic domains and semantic tagging. Each of the four sub corpora of the node words alcohol, tradition, orange, and tea—was fed to Wmatrix individually. After passing through corpus analysis tools a list of the ninety to hundred semantic fields from the elicited data was generated along with semantic tagging.

The sub-semantic corpora's domains (elementary meaning units) are listed below (tradition, alcohol, orange, and tea)

Table 2: Semantic domains in tradition corpus of the elicited data

Semantic domains	**Semantic tagging**	**Semantic domains**	**Semantic tagging**	**Semantic domains**	**Semantic tagging**
Social actions	S1.1.1	Informal/friendly	S1.2.1+	Time: old, new, and young; age	T3
Food	F1	Quantities: much	N5++	Vehicle and transport on land	M3
Unmatched	Z99	Degree: compromisers	A13.5	Easy	A12+
Drink	F2	Conceptual objects	X4.1	Residence	H4

Below are a few examples from the "tradition" corpus.

Semantic Domains	**Concordances**
Food	Dinner is generally served with **breads** like **roti or naan**. Popular lunchtime fare might include **okra aloo gosht, beef and potato**. Traditional **foods** include **poultry meals like chicken karahi.**
Drinks	The traditional **drinks** in summer are **Jam e Shireen, lassi** and **brown tea**. The traditional **drink** in winter is **green and black tea**

Clothes	**Veils** and **headwear** for women are part of the tradition.
Parts	Holding the **Quran** on the head of the bride is also **part** of our tradition.

Nearly 90% of the elementary meaning units in the "Tradition" corpus are retrieved using semantic domains. Semantic domains are the higher-level, wider categories that encompass related types and tokens. By comparing the node words in each semantic domain, one may determine the basic meaning units. EMUs (elementary meaning units), as was already said, are a person's subjective response to a certain (node) word.

Table 3: Semantic domains of alcohol corpus

Semantic domains	**Semantic domains**		**Semantic domains**		
	Semantic Tagging		**Semantic tagging**		**Semantic tagging**
Drinks/alcohol	F2	Anatomy/physiology	B1	Comparing: Usual	A6.2+
Not allowed	S7.4-	Quantities: much/many	N5++	Violent/Angry	E3-
Religion/supernatural	S9	Evaluation: bad	A5.1-	Grammar	Z5
Using	A1.5.1	Sensible	S1.2.6+	Change	A2.1+

About 98 per cent of the semantic domains and semantic tagging are created in the second sub corpus, "Alcohol." The following instances of concordance for the node word "alcohol" further build on the identification of elementary meaning units (EMUs) via semantic domains to determine their meaning.

Semantic domains	**Concordance**
Drink	It is forbidden for Muslims in Pakistan to drink alcohol, and discussing it is frowned upon.
Not allowed	Alcohol is **banned, and forbidden** in Islam, according to hadith it is the root of all illness
Religion	**Pious** people avoid its use even when medically required as there exists some relaxation in **Islam/religion**

Geographical names

While alcohol consumption is restricted in Khyber Pakhtunkhwa KP some individuals do consume it privately

Table 4: Semantic domains of orange corpus

SEMANTIC DOMAINS	SEMANTIC TAGGING	SEMANTIC DOMAINS	SEMANTIC TAGGING
Food	F1	Language /speech	Q3
Color/color patterns	O4.3	Health and disease	B2
Time: new and young	T3---	Communication	Q1.1
Possession	A9	Existing	A3+
Degree: booster	A13.3	Exclusion	A1.8-

Semantic domains	Concordances
Food	I enjoy orange **marmalade** on my **toast**
Color/patterns	Orange **shade** lipsticks suit a lighter cool **tone** of skin
Showings, findings	Orange is an **identity** of Hindu Orange **reflects** the ideology of BJPs in India
Drink	A course of orange **Juice** roast chicken and a cup of tea was preferred

Table 5: Semantic domains of “tea” corpus

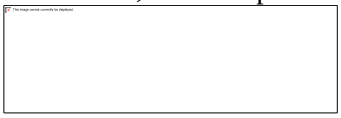
Semantic domains	Semantic tagging	Semantic domains	Semantic tagging	Semantic domains	Semantic tagging
Drink	F2	Calm	E3+	Temperature: cold	O4.6-
Color/ pattern	O4.3	Smoking	F3	Mental action	X2
Using	A1.5.1	Evaluation: Good	A5.1+	Food	F1
Like	E2++	Farming and horticulture	F4	Emotional action	E1

To extract Elementary Meaning Units (EMUs) in the form of semantic domains, the node word "tea" was included in the study. Through the use of the Wmatrix corpus analysis tool, the sub corpus "tea" generates around 95% of the total semantic domains. The examples that illustrate the semantic domains using concordance are listed below.

Semantic domains	Concordance
Color/pattern	**Green** tea is served when eaten meat
Substance: generally	Tea contains **caffeine** and **nicotine**
Furniture	Brewed tea is used to clean wood floors and **furniture**

4.2 Level of Conventionalization of Semantic Fields

For the analysis of question no. 2 in the dissertation which is to find the level of conventionalization of the respective semantic domains Molinari's evenness measure ($G_{2,}$) was chosen. Beisel et al. (2003) thoroughly evaluate and contrast many evenness indexes that are currently available. Molinari's evenness index is deemed appropriate for this study since it assigns higher weight than other indices to high, medium, and low conventionalization levels of the corresponding semantic areas. Molinari's (1989) evenness index ($G_{2,}$) is calculated on Hill's (1973) modified evenness, which is provided as follows (F_2)

$1F_{2,1}$ 

To acquire $G_{2,1}$ we need to observe ($F_{2,1}$) and use the weightings below;

Molinari's evenness index measure was thought to be pertinent for determining the degrees of conventionalization of the various semantic domains in the elicited data since culture is a group construct made up of collective views, ideologies, and traditions. To extract the total number of tokens from each semantic domain of the sub-corpora

Wmatrix 5 was used to determine the extent of conventionalization using Molinari's evenness metric.
"Molinari's evenness measure (G2)" was used to evaluate the degree of conventionalization within each semantic field, drawing inspiration from Wilson and Mudraya (2006). High evenness occurs in a community where all species are equally plentiful, whereas low evenness occurs in a community where species abundances range greatly". The current study viewed each semantic field as a "community/region" and each subject as a "species / taxon" that appears (in that semantic field) a given number of times, contributing a defined numeral occurrence to the community's composition. In the context of this research, Molinari's G2 indicates that it is most responsive to small fluctuations in dominant that is high level and medium semantic domains and relatively sensitive to changes in uncommon semantic domains mainly low and non-
Conventionalized levels of the semantic domains. This suggests that it is highly sensitive to even little variations in the frequency of a certain semantic region in each respondent's replies. Using raw counts, the Molinari index is constructed. The evenness values are then separated into three categories, as in the (Wilson and Mudraya 2006) study, which paralleled "high (H), medium (M), and low (L)" degrees of conventionalization.

4.3 Tradition

Based on the semantic fields extracted from sub-corpora in analysis no.1 further results are obtained using Molinari's evenness measure considering the semantic domains as a region and calculating the number of types and tokens in each of the tradition, alcohol, orange, and tea sub corpora.

4.4 Tradition

Table 6: Level of conventionalization of the "tradition" corpus

Semantic domains	**$F_{2,1}$**	**G2**	**Level of Cnvt.**	**Semantic Domains**	**$F_{2,1}$**	**G2**	**Level of Cnvt**
Social actions	0.0102	0.0038	H	Kin	0.0634	0.1439	M
Music	0.1540	1.00	L	Not allowed	Not valid	Not valid	NC

Table7: level of conventionalization of "Alcohol" corpus

Semantic domains	**Fo**	**G2**	**Level of Cnv**	**Semantic domains**	**Fo**	**G2**	**Level of cnv.**
Drink	0.0083	0.0027	H	Excessive drinking	0.0598	0.1117	M
Politics	0.1615	0.9906	L	Helping	Not valid	Not valid	NC

Table 8: level of conventionalization of "orange "corpus

Semantic domains	**Fo**	**G2**	**Level of Cnv.**	**Semantic domains**	**Fo**	**G2**	**Level of Cnv.**
Food	0.0127	0.0057	H	Like	0.0598	0.1348	M

Content	0.0360	1.00	L	Media	Not valid	Not valid	NC

Table 9: level of conventionalization of "tea" corpus

Semantic domains	Fo	G2	Level of Cnv.	Semantic domains	Fo	G2	Level of Cnv.
Color/pattern	0.0267	0.0202	H	Medicine/treatment	0.0527	0.1061	M
Farming/horticulture	0.1031	1.00	L	Shape	Not valid	Not valid	NC

The tables (4.2.1, 4.2.2., 4.2.3, 4.2.4) show semantic domains, Fo (Hills evenness indices), G2 "Molinari's evenness measures", and the level of conventionalization of the required tradition, alcohol, orange, and tea corpora. H stands for a "high level of conventionalization" which means the semantic domain in the respective sub-corpora (tradition, alcohol, orange, and tea) is more prominent than the other semantic fields. M stands for "medium level of conventionalization" and L denotes "low level of conventionalization". To overcome the lack of correspondence in the conventionalization values, the analyses are carried out by rounding to the fourth decimal place. There are situations when rounding to two decimal places results in the same degree of conventionalization. Such instances of seeming incoherency are quite uncommon, though. There were less than two occurrences in that field, the "Molinari's evenness formula" was unable to derive and calculate value, which means that the frequency of that particular semantic field is one, and is indicated by the NC in the evenness column which means it is not conventionalized as it is a personal opinion rather than cultural trait.

The results of Molinari's evenness index are inversely correlated with the degree of conventionalization. The conventionalization level of the tradition corpus decreases as the evenness index values rise. Semantic domains such as social actions, foods, drinks, clothes, and entertainment have a high level of conventionalization, as illustrated in Figure 4.2.1. On the other hand, beliefs, kin, arts and crafts, and belonging to a group have a medium level of conventionalization, while media, bravery, residence, and helping have a higher evenness measure, indicating a low level of conventionalization. Similarly, a corpus of "Alcohol" from
0.0005-0.0942 in Figure 4.2.2 displays high conventionalization values, medium levels of conventionalization from 0.1117-0.5728, and lower levels of conventionalization from 0.9906-
1.00 evenness index.

The degree of conventionalization varies among the "orange and tea" corpora shown in Figures 4.2.3 and 4.2.4, on the other hand. A higher level of conventionalization of important semantic domains in the culture is indicated by evenness values of 0.0002-0.0720 and 0.0202-0.0986; a medium level of conventionalization is indicated by values of 0.1063-0.4921 and 0.1061-0.7990; a lower level of conventionalization is indicated by 1.00 of the corpora respectively.

High conventionalization fields often concentrate in the top rankings in four of the tables. On the other hand, low conventionalization fields start to show up about midway down the list and get more concentrated towards the end. This appears to imply that fields holding NC values might be regarded as having a low degree of conventionalization. Consequently, fields marked with NC show "not conventionalized" indicating less than two occurrences in the semantic fields. This
NC (not conventionalized) is the hapax legomena to which "Molinari's evenness index" does not apply.

4.5 *Analyzing* culture through theoretical framework

For the analysis of the third objective of the study, a theoretical cultural framework of Fleischer is taken.

"Fleischer distinguishes three components that comprise social and discursive representations: core, present field, and connotation field. One stable component is the core. Enduring collective signs have a central, powerful element. The present field is an aspect that is still developing and is somewhat generic. Cultural connotations are expressed in both present and core fields. Lastly, the connotation domain is linked to the specific lexical meaning

and is a representation of a single meaning; consequently, it has no stabilization at all."
Based on the study of the second question, Fleisher's cultural idea is used to simplify semantic fields with varying degrees of conventionalization and emphasize the key semantic domains in Pakistani culture.
The following table shows a summary of the levels of conventionalization of semantic fields in terms of the percentage of each sub-corpora (tradition, alcohol, orange, and tea).

Table 10: Percentages of sub-corpora 2

Corpora	**H (core)**	**M-L (current)**	**NC (connotation)**
Tradition	55%-60%	37%	3%
Alcohol	40%-46%	51%	3%
Orange	85%-90%	12%	0.5%
Tea	80%-87%	11%	[1]%

The third objective of the study analyzed through Fleischer's cultural theoretical framework allows us to comprehend and clarify the traits of a particular culture through core, current and connotation elements. The higher conventionalization level of elementary meaning units(EMUs) in the form of semantic domains of particular node words (tradition, alcohol, orange and tea) show that these semantic fields are the core elements of the culture and are ingrained in the culture which every individual in that society shares and are very stable elements whereas, the medium to low level of conventionalization of semantic domains depict the current element of Fleischer cultural concept which shows these elements are generalized but not shared by everyone in the community. Connotation fields are individual opinions and do not exist in the culture they are rather individual opinions and are highly unstable elements.

5. Conclusion and Recommendations

The present probe is about understanding a Pakistani culture utilizing a corpus analytical tool like Wmatrix 5, through elicited data i.e. data collected from the participants. A corpus of 35,023 words (node words: tradition, alcohol, orange, and tea) of data was extracted from 158 (one hundred and fifty-eight) respondents utilizing the concept of Elementary meaning units which is a subjective reaction to individual words. The individual words called node words were tradition, alcohol, orange, and tea. After the formation of the sub-corpora (tradition, alcohol, orange, and tea) these were fed to Wmatrix 5 and wordsmith corpus analysis tools. About 90%95% of the EMUs were extracted in the form of semantic fields and semantic tagging (figure 4.1.1, 4.1.2. 4.1.3, and 4.1.4) of each tradition, alcohol, orange, and tea along with concordances. After retrieving the semantic domains through corpus analysis tools the next step was to find the levels of conventionalization for each domain of each sub-corpora. The method employed was through Molinari's evenness measure. High level, medium land low levels of conventionalization were found through Molinari's evenness index. Through analyses, it was found that the levels of conventionalization are inversely proportional to Molinari's evenness values (figure 4.2.1, 4.2.2. 4.2.3 and 4.2.4). The higher the evenness values lower the level of conventionalization of semantic domains of each sub-corpora and vice-versa. It was found that Molinari's evenness index does not apply to hapax legomena (a single occurrence of an entity at a time) and is denoted by NC (not conventionalized). Through the level of

[1] The percentages estimated based on the number of times it occurred i.e. frequencies of level of conventionalization of semantic domains of each sub-corpus.

conventionalization, it is evident which semantic fields are dominant in a culture. Semantic fields were further analyzed through

Fleischer's cultural concept. Fleischer (1998) created a systemic model of culture that is accessible to both quantitative and semantic analysis to broaden knowledge of cultural systems and the connections between culture, language, and semantics particularly semantic fields. . About 55%-60% of tradition, 40%-46% of alcohol, 85%-90% of orange, 80%-87% of tea semantic fields are the core element of the culture. Medium to low level of conventionalization of semantic fields form the current element of the culture which is not much stable and not shared by all members of community.37% tradition, 51%alcohol, 12% orange, and 11% tea forms the current field of culture whereas 3% tradition and alcohol, 0.5% orange and 2% tea of total semantic fields form the connotation element and is not conventionalized which means it is an individual opinion and does not fit in to the collective concept of the culture. Thus, Wmatrix corpus analytical tool, levels of conventionalization and Fleischer concept can be used for marketing research, social media research and for cultural studies.

References

Bianchi, F. (2010). Understanding Culture. Automatic Semantic Analysis of a General Web Corpus and a Corpus of Elicited Data. ETC, 4, 1-29

Fink, A. (2003). The survey handbook. Sage.

Fleischer, M. (2002). Das Image von Getränken in der polnischen, deutschen und französischen. Kultur,[w:]. Empirische Text-und Kulturforschung, 8-47

Greenberg, J. H. (1966). Language universals . The Hague: Mouton., 9-32.

Hair, J. F., Bush, R. P., & Ortinau, D. J. (2009). Marketing research in a digital environment.

Hall, E. T. (1989). Beyond culture. In: Doubleday.

Halliday, M. A. K. (1978). Language as social semiotic: The social interpretation of language and meaning. (No Title).

Howell, K. E. (2013). Empiricism, positivism and post-positivism. An introduction to the Philosophy of methodology, 32-54.

Phillips, D. C., & Burbules, N. C. (2000). Post positivism and educational research: Rowman

Ryan, G. (2018). Introduction to positivism, interpretivism and critical theory. Nurse Researcher, 25(4), 41-49.

Shuttleworth, W. J. (2008). Evapotranspiration measurement methods. Southwest Hydrology, 7(1), 22-23.

Szalay, L. B. and Maday,B.C.,(1973). Verbal Associations in the Analysis of Subjective Culture. Current anthropology,, 33-50.

Williams, R. (1976). Keywords: A Vocabulary of Culture and Society (New York, 1976). WilliamsKey Words: A Vocabulary of Culture and Society1976.

Wilson, A., & Mudraya, O. (2006). Applying an evenness index in quantitative studies of language and culture: a case study of women's shoe styles in contemporary Russia.

Appendices

Tables

Semantic Extraction of Sub-Corpora through Wmatrix.

Table 11: Semantic Domains in Tradition Corpus of the elicited data

Semantic domains	**Semantic tagging**	**Semantic domains**	**Semantic tagging**	**Semantic domains**	**Semantic tagging**

Social actions	S1.1.1	Informal/friendly	S1.2.1+		T3
				Time: old, new, and young; age	
Food	F1	Quantities: much	N5++		M3
				Vehicle and transport on land	
Unmatched	Z99		A13.5	Easy	A12+
		Degree: compromisers			
Drink	F2	Conceptual objects	X4.1	Residence	H4
Happy	E4.1+	Time: new/young	T3--	Helping	S8+
Religion	S9	No cautions	A1.3-	Comparing: similar	A6.1+
People	S2	Children games	K6	Tough/strong	S1.2.5+
Entertainment	K1	Long, tall, wide	N3.7++	Bravery	E5+
Kin	S4	Speed: slow	N3.8-		A4.1
				Generally, kind, group examples	
Existing	A3+	knowledgeable	X2.2+	Psychological action	X1

Tasty	X3.1+	Time: present; simultaneous	T1.1.2	The media	Q4
Clothes	B5	Time: old/grow nup	T3++	Weather	W4

Degree: maximizers	A13.2	Degree: non-specific	A13.1	Seem	A8

Geographical names	Z2	Infrequent	N6-	Belongings to a group	S5+
Arts and crafts	C1	Language; Speech	Q3	Mental objects: means and methods	X4.2
Like	E2+	Like	E2+++	Law and order	G2.1
Comparing: Usual	A6.2+	Texture	O4.5	Thought, beliefs	X2.1
Comparing: similar/different	A6.1	Part	N5.1-	Comparing: varied	A6.3+
Sensory: taste	X3.1	Color/pattern	O4.3	Not allowed	S7.4-
People: Female	S2.1	Quantities: Little	N5---	Short and narrow	N3.7-
Cause and effect connection	A2.2	Games	K5.2	General Ethics	G2.2
Respect	S7.2+	Lawful	G2.1+	Size: big	N3.2++

Important	A11.1+	Cleaning and personal care	B4	Flying and aircrafts	M5
Places	M7	Personality traits	S1.2	Time: Early	T4+
Farming and horticulture	F4	Substance and materials; liquids	O1.2	No power	S7.1-
Sports	K5.1	The Universe	W1	Alive	L1+
Inclusions	A1.8+	Government	G1.1	Xx(ancestry)	S4T1.1.1
Period	T1.3	Evaluation: bad	A5.1-		
Change	A2.1+	Ethical	G2.2+		
Money: affluence	I1.1+	Music	K2		

Table 12: Semantic domains of alcohol corpus

Semantic domains	**Semantic Tagging**	**Semantic domains**	**Semantic tagging**	**Semantic domains**	**Semantic tagging**
Drinks/alcohol	F2	Anatomy/physiology	B1	Comparing: Usual	A6.2+
Not allowed	S7.4-	Quantities: much/many	N5++	Violent/Angry	E3-
Religion/supernatural	S9	Evaluation: bad	A5.1-	Grammar	Z5

Using	A1.5.1	Sensible	S1.2.6+	Change	A2.1+
Cause and effect	A2.2	Exceed: waste	N5.2+	Expensive	I1.3+
Health/disease	B2	Deciding	X6	Safety/danger	A15
Damaging/destroying	A1.1.2	Interesting/energetic	X5.2++	Politics	G1.2
Tough/strong	S1.2.5+	General substance	O1	Constraint	A1.7+
Medicine/treatment	B3	Hindering	S8-	Period: long	T1.3+
Calm	E3+	The media: TV/Radio	Q4.3	Professional	I3.2+

Danger	A15-	Time: old/grown-up	T3+++	Non-participating	S1.1.3-
Unethical	G2.2-	Speech acts	Q2.2	Alive	L1+
People	S2	Business: selling	I2.2	The Universe	W1
Conceptual object	X4.1	Cleaning/ personal care	B4	Important	A11.1+++
Important	A11.1+	Sound: quiet	X3.2-	Size: Small	N3.2-

Social actions	S1.1.1	Degree: maximizers	A13.2	IT/computing	Y2	
Avoiding	A1.9	Open/findings/showing	A10+	Law and order		G2.1
Ethical	G2.2+	Comparing: different	A6.1-	No power	S7.1-	
Dislikes	E2-	Time: late	T4--		N3.7++	
				Long, tall and wide		
Diseases	B2-	Arts and crafts	T4--	Weight: Heavy	N3.5+	
Geographical names	Z2	Smoking/ drugs	F3	Speed: slow	N3.8-	
Mental actions	X2	Psychological actions	X1	Difficult	A12-	
Success	X9.2+	Infrequent	N6-	Plants	L3	
Inclusion	A1.8+	Substance: gas	O1.3	Worry	E6-	
Unmatched	Z99	The media	Q4	Wanted	X7+	
Crime	G2.1-	Comparing: varied	A6.3+			N3.3
				Measurement/distance		
Government	G1.1	Emotions: happy	E4.1+	Thoughts/belief	X2.1	

Language/speech	Q3	Easy	A12+	Money: debts	I1.2
Quantities	N5	Frequent	N6+		S6+
				Obligations/necessities	
Excessive drinking	F2++	Measurement: speed	N3.8	Emotional actions	E1
Trying hard	X8+	Attentive	X5.1+		A13.1
				Degree: non-specific	
Evaluation: true	A5.2+	Cautious	A1.3+	Kinds, examples	A4.1

Table 13: Semantic domains of orange corpus

SEMANTIC DOMAINS	SEMANTIC TAGGING	SEMANTIC DOMAINS	SEMANTIC TAGGING
Food	F1	Language /speech	Q3
Color/color patterns	O4.3	Health and disease	B2
Time: new and young	T3---	Communication	Q1.1
Possession	A9	Existing	A3+
Drink	F2	Comparing: Usual	A6.2+
Likes	E2++	Business: Generally	I2.1
Easy	A12+	Helping	S8+
Showings	A10+	Location and direction	M6
Quantities: much/many	N5+	Drama, theatre /show	K4
Substance and material general	O1	Moving: coming/going	M1

Giving	A9-	Linear order	N4
Degree: booster	A13.3	Exclusion	A1.8-
Tasty	X3.1+	Clothes and personal belongings	B5
Objects: generally	O2	Healthy	B2+
Sensory: taste	X3.1	Temperature: hot	O4.6+
Like	E2+++	Expensive	I1.3+
Personality: traits	S1.2	Emotional actions	E1
Anatomy/physiology	B1	Distance: near	N3.3---
Light	W2	Content	E4.2+
Cleaning and personal care	B4	Expected	X2.6+
Foolish	S1.2.6-	Sensory: smell	X3.5
Money: Affluence	I1.1+	Evaluation	A5.1+

Table 14: Semantic domains of "tea" corpus

Semantic domains	**Semantic tagging**	**Semantic domains**	**Semantic tagging**	**Semantic domains**	**Semantic tagging**
Drink	F2	Calm	E3+	Temperature: cold	O4.6-
Color/ pattern	O4.3	Smoking	F3	Mental action	X2
Using	A1.5.1	Evaluation: Good	A5.1+	Food	F1
Like	E2++	Farming and horticulture	F4	Emotional action	E1

Exclusion	A1.8-	Getting and possession	A9+	Tough/strong	S1.2.5+
Helping	S8+	Healthy	B2+	Measurement: weight	N3.5
Darkness	W2-	Disease	B2-	No constraints	A1.7-
Anatomy/ physiology	B1	Personal care	B4	Frequent	N6+++
Existing	A3+	Time: Beginning	T2+++	Sensory: smell	X3.5+
Health and disease	B2	People	S2	Comparing: different	A6.1-
Inclusion	A1.8+	Though, beliefs	X2.1	Temperature: hot	O4.6+
Substance/generally	O1	Substance: solid	O1.1	No obligations	S6-
Sensory: taste	X3.1	Personal relation	S3.1	Residence	H4
Medicine/ treatment	B3	Comparing: Usual	A6.2+	Damaging/destroying	A1.1.2
Drama and show	K4	Danger	A15-	Quantities: much	N5+++
Object generally	O2	Foolish	S1.2.6-	Money: affluence	I1.1+

Texture	O4.5	Tasty	X3.1+	Lack of food	F1-
		Frequency	N6	Plants	L3
Expensive/unenergetic	I1.3+/ X5.2-				
Quantities: Little	N5---	Psychological actions	X1	Unmatched	Z99
Arts and crafts	C1	Parts	N5.1-	Exceed ;waste	N5.2+
Uninterested; excited	X5.2+	Weight; heavy	N3.5+	Happy	E4.1+
Worry	E6-	Degree; minimizers	A13.7	Size; big	N3.2+
Likely	A7+	Dislikes	E2-	Hindering	S8-
judgement	O4.2-	Quantity; little	N5-	Avoiding	A1.9
Relationship, intimacy, sex	S3.2	Shape	O4.4	Crime	G2.1-

Table 15: Level of conventionalization of the “tradition” corpus
Tradition

Semantic domains	**Fo**	**G2**	**Level of Cnvt.**	**Semantic Domains**	**Fo**	**G2**	**Level of Cnvt**
Social actions	0.0102	0.0038	H	Kin	0.0634	0.1439	M

Clothes	0.0467	0.0686	H	Respect	0.1341	0.5096	M
Religion	0.0193	0.0042	H	Though/beliefs	0.0692	0.1569	M
Entertainments	0.0480	0.0733	H	Change	0.0545	0.1091	M
Colors/patterns	0.0440	0.0675	H	Residence	0.1787	1.000	L
Cleaning/ care	0.0311	0.0355	H	Helping	0.2244	1.000	L
Conceptual objects	0.0422	0.0675	H	Bravery	0.2509	1.000	L
Degree/compromisers	0.0440	0.0675	H	Media	0.1641	1.000	L
Money/affluence	0.0346	0.0395	H	Music	0.1540	1.00	L
Vehicles: land	0.0281	0.0213	H	Quantity: much	Not valid	Not valid	NC
Parts	0.0167	0.0063	H	Psych. Actions	Not valid	Not valid	NC
Means and methods	0.0394	0.0446	H	Not allowed	Not valid	Not valid	NC
Games	0.0894	0.2861	M	Politics	Not valid	Not valid	NC
Lawful	0.1161	0.4879	M	Inclusion	Not valid	Not valid	NC
Arts and crafts	0.1285	0.5828	M				

Table 16: level of *conventionalization* of "Alcohol" corpus

Semantic domains	**Fo**	**G2**	**Level of Cnv**	**Semantic domains**	**Fo**	**G2**	**Level of cnv.**

Drink	0.0083	0.0027	H		0.0598	0.1117	M
				Excessive drinking			
Not allowed	0.0147	0.0078		Open, findings	0.1227	0.5728	M
			H				
Religion	0.0192	0.0141	H	Business	0.1063	0.4041	M
Heath/disease	0.0381	0.0005	H	Media	0.0598	0.1117	M
Unethical	0.0432	0.0720	H	Politics	0.1615	0.9906	L
Strong/tough	0.0259	0.0335	H	Kind, examples	0.2261	1.0000	L
Obligation	0.0642	0.0128	H	Worry	0.1641	1.0000	L
Anatomy/physiology	0.0357	0.0405	H	Inclusion	0.2189	1.000	L
Geographical names	0.0308	0.0345	H	Helping	Not valid	Not valid	NC
Medicines	0.0504	0.0942	H	Quantity: much	Not valid	Not valid	NC
Crimes	0.0312	0.0457	H	Food	Not valid	Not valid	NC

Social actions	0.1463	0.5560	M	Clothes	Not valid	Not valid	NC
Government	0.0972	0.3605	M				

Table 17: level of conventionalization of "orange "corpus

Semantic domains	**Fo**	**G2**	**Level of Cnv.**	**Semantic domains**	**Fo**	**G2**	**Level of Cnv.**
Color/pattern	0.0267	0.0202	H	Medicine/treatment	0.0527	0.1061	M
Substance: generally	0.0481	0.0808	H	Relationship/intimacy	0.0929	0.3251	M
	0.0342	0.0444	H	Furniture	0.0625	0.1431	M
Geographical names	0.0447	0.0238	H	Tough/strong	0.0460	0.7990	M
Smoking/medical drugs	0.0395	0.0587	H	Lack of food	0.0804	0.2471	M
People	0.0289	0.0317	H	Plants	0.0658	0.1623	M