Journal of Linguistic and Rhetorical Studies
Volume 17, Consecutive Number 43, Spring 2026
Pages 107-134 **(research article)**
Received: June 05, 2025 **Revised** July 13, 2025 **Accepted** July 22, 2025
Journal Homepage: https://rhetorical.semnan.ac.ir/?lang=en

Bidirectional Interaction of Psycholinguistics and Corpus Linguistics; Some Challenges Persian Frequency Dictionaries Meet

Masoumeh Mehrabi [1*] **& Behrooz Mahmoodi-Bakhtiari** [2]
1: Associate Professor of Linguistics at Ayatollah Boroujerdi University, Broujerd, Iran (Corresponding Author): m.mehrabi@abru.ac.ir
2: Associate Professor of Linguistics at Fine Art College of Tehran University, Tehran. Iran.

Abstract: Lexical decision tasks in psycholinguistics is the exact intersection of psycholinguistics and corpus linguistic findings and outputs. Here is the ground which the corpus analyses will be used directly. Throughout the course of using the frequency dictionaries of Persians the researcher(s) may encounter(s) some shortcomings and deficiencies that lead to the conclusions: Frequency estimates are not sufficient. Part of this relates to the fact that the word frequency as depicted in frequency dictionaries is affected by other factors (like the kind of sources used in data compilation) and the other part relates to the effects of other variables which correlates with word knowledge (like word prevalence, word length, age of acquisition of the word, similarity to other words, and concreteness of the word content). Here, the research problem is that it seems current Persian frequency dictionaries are insufficient for psycholinguistic researches. The present article tries to find why it is the case. The main question of the present study is the psychological reality of the frequency effects which two main Persian frequency dictionaries (abbreviated as RPFD and PFD here) are able to depict. It asks what psycholinguistic factors will provide corpus compilation methodology with frequency dictionaries to be more realistic and valid. Using lexical decision tasks, the present article shows that some other psychological factors (like the recency effect, prominence, and prevalence) will correlates with word knowledge which these dictionaries miss to show or they ignore. Considering these factors will enhance Persian frequency dictionaries. The article systematically addresses each limitation (e.g., lack of prevalence data, recency effects). It also proposes solutions (e.g., integrating psycholinguistic metrics into corpus design).

Keywords: Corpus Linguistics, Frequency Dictionary, Psycholinguistics, Persian, Frequency of occurrence.

- M. Mehrabi; B. Mahmoodi-Bakhtiari (2026). "Bidirectional Interaction of Psycholinguistics and Corpus Linguistics; Some Challenges Persian Frequency Dictionaries Meet". Semnan University: *The Journal of Linguistic and Rhetorical Studies* 17(43). 107-134.
Doi: 10.22075/jlrs.2025.37997.2679

1. Introduction

Psycholinguistics, particularly the branch that employs experimental methodology, based on laboratory tests overlaps cognitive linguistics. Here, we have confined the scope of the research to examining how corpora have been used in some psycholinguistic investigations of language processing mainly through lexical decision tasks. Since frequency has certain effects on the processing of linguistic items and ultimately the language system itself psycholinguistic investigations are increasingly making use of corpus linguistics information about frequency. Frequency effects which are ubiquitous in virtually every domain of human cognition influence these cognitive tests and tasks directly in two ways; corpus data can be used as a check on naturalness of the language tasks/tests and they can be used as the source of frequency data in the construction of the test sentences/ items, too.

The main questions of the recent investigation are:

Why is it the case that the interaction of psycholinguistics and corpus linguistics seems to be in large degree a one-way street in favor of psycholinguistics? Why are much of the relevant researches which have used corpora "psychologically-oriented" rather than "corpus- linguistically oriented"? What are the deep deficiencies of Persian frequency dictionaries and why is the application of experimental methods of psycholinguistics in corpus samples so limited?

The hypothesis is that answers to the above questions will help the methodologies of frequency dictionaries to be much more standardized.

This is a corpus-driven study since the corpus is the source of the hypothesis about language processing.

Two main frequency dictionaries of Persian (*Routledge Frequency dictionary of Persian -abbreviated by RPFD, here* and the *frequency dictionary according to a written corpus of today Persian language -abbreviated by PFD, here*) are the main sources of investigation whose methodology of corpus construction and analysis will be investigated, discussed, and the shortcomings will be depicted and referred clearly.

As for the deficiencies of contemporary frequency dictionaries of Persian (discussed in the article body in detail) much of the psycholinguistic investigations are corpus-assisted rather than

corpus- based. Resolving this problem will help psycholinguistic investigations to be elevated to corpus- based ones. Besides, there will be some suggestions for the methodologies of frequency dictionaries to be much more standardized so that there can be found some concepts like word prevalence, salience, entrenchment, lexical collocation, and automatization which are cognitive concepts beside frequency effects. Employing such psychological terms like these in frequency dictionaries will lead this type of dictionaries to be much more cognitive-oriented ones in accordance with the human cognitive apparatus. In this way the contribution of cognitive psychology and corpus linguistics will be balanced in the frame and formation of new frequency dictionaries.

2. Literature Review

2.1. Corpus Linguistics

The history of corpus linguistics may be divided into two main phases:

1. The first phase, which lasted until the end of the 1980s, focused on the development of corpus linguistics at two distinct English language study schools:

a. Its battle to gain traction against the Chomskyan viewpoint, which is fundamentally hostile to the usage of corpora.

b. The creation of the foundational set of techniques and equipment for corpus compilation.

2. From that time onward, the central subject of corpus linguistics has been the change in the field's characteristics, from being a practically autonomous subfield of linguistics to an essential part of the methodological toolkit via linguistics.

This status—where corpus linguistics approaches will be employed by all linguists instead of serving as the guardian of a marginalized subset, as was the case up until the 1990s—may be the future progression projection.

2.2. Corpus Linguistics and Psycholinguistics

Cognitive linguistics, in particular the branch whose technique is mostly experimental, intersects with the broader topic of psycholinguistics. In this case, several human subjects' laboratory tests serve as the main source of data. Generally, the use of corpus-based approach and findings is becoming more common in the field of experimental psycholinguistics. Psycholinguistic studies and other corpus-driven methods reject theoretical frameworks that are

not driven by a corpus. By emphasizing categories derived from the data itself, corpus linguistics appears to provide a viewpoint on language that differs from the accepted ideas of cognitive linguistics.

Psycholinguistics is a fairly large area, therefore we will limit our discussion to a very quick examination of one of its methodologies, the lexical decision, and even though there are many ways in which language processing has been studied experimentally.

Corpus data may be useful in two ways for the planning and analysis of these kinds of experiments:

Firstly, the naturalness of the language task that participants in the experiment are given may be verified using corpus data.

Secondly, in the lexical decision experiments, test sentences may be constructed using frequency statistics obtained from corpus data.

Researchers were left with limited options for a long time when it came to word frequency measures. Only one or two lists, if any, existed for each language due to the time-consuming nature of word counting in printed books and newspapers. When texts were available digitally, things drastically altered. After then, compiling a corpus—a sample of texts—and counting the words inside them got considerably simpler. Word frequencies mostly drawn from nonfiction publications like scientific books, periodicals, and newspapers. Linguistic corpora are widely used in linguistic researches like those of Zolfaghari and Hasini Kochaki (2023).

The most effective word frequency measures for undergraduate students—the group most frequently involved in psychology studies—were found to be based on corpora of television subtitles (Brysbaert & New, 2009), social media (Gimenes & New, 2016; Herdag˘delen & Marelli, 2017), and blogs (Gimenes & New, 2016). Generally speaking, combining various sources yields superior outcomes than using them separately. Traditional word frequency measurements based on books are sometimes more appropriate for older individuals (Brysbaert & Ellis, 2016). According to Johns, Jones, and Mewhort (2016), creating frequency lists specific to study participants based on their learning histories—that is, what kind of television they watch, what authors they read, how active they are on social media, and what textbooks they use—might yield additional benefits. Well-crafted frequency lists rely on a substantial corpus, with no fewer than 20 million words. These lists also include

information on the syntactic functions that words (verbs, for example) fulfil, allowing researchers to utilize this data as well. Furthermore, effective frequency lists must to have demonstrated their worth in validation experiments using data from mega studies (Brysbaert, Mandera, and Keuleers, 2018).

3. Method

3.1. The Need of Psycholinguistics for Frequency Dictionaries

Why does frequency matter? The number of times an object, event, or item has been encountered is known as its frequency, and frequency effects are pervasive in all areas of human cognition. Experience is important for our brain machinery. Encounters matter cognitively. Processing language is not an exception to this rule. For language learners, speakers, and listeners, language experience is significant. Word frequency, or the frequency with which a linguistic item occurs, is thought to be a predictor of both processing speed and efficiency.

One of the most widely used features of words is frequency. When lexicographers are determining whether or not to add a word to a dictionary, this is frequently their primary consideration. The number of times a word appears in a text is its frequency, and this is straightforward to compute. We can quickly determine which terms are common by comparing the frequencies of all the words in a particular text. That outcome, however, solely relates to the text that we utilized in our computations—it does not address the language in its whole. The text determines the frequency in terms of word occurrences, not just on how long it is, but also on its topic, author(s), style, and other characteristics.

However, what does the word "frequency effects" actually mean? According to the word frequency effect (Monsell, Doyle, & Haggard, 1989), high-frequency words are processed more quickly and are known to a larger audience than low-frequency terms. This is true for tasks like word naming, semantic decision-making (e.g., does the term relate to an animal?), and lexical decision-making (does the letter string belong to an existing word or not?). Word frequency has an impact on memory function as well. In memory studies, participants examine a list of words before being asked to recall them or tell them apart from lures (new things). It's interesting to note that the task affects the pattern of results: Overall, low-

frequency words are harder to remember yet perform better on recognition tasks (Yonelinas, 2002).

While it is not maintained that frequency is the most important component in language, we also do not want to claim that the most frequent item(s) drives all processes. Type and/or token frequencies have a direct impact in some domains, while frequency interacts with other processing parameters like salience and recency in other constellations. For example, automatization and entrenchment result from frequency in terms of high transition probability. Because of its entrenchment, high frequency can guard against mistakes in some forms, but it can also create mistakes in situations when a lower frequency form is the intended goal (Ambridge et al., 2015).

Niemikorpi (1997) believed that frequent words behave in such a way that make them distinct from other words:

1. High-frequency words often include common and mundane words of the language, while low-frequency words are mainly words that have a short life in the language or they are or words that have recently become popular.
2. High-frequency words are often composed of a single morpheme, while low-frequency words usually consist of several morphemes.
3. High-frequency words are usually more irregular than low-frequency words from the point of view of word formation and morphology.
4. High-frequency words have a more general meaning from the semantic point of view, so the number of their semantic distinguishing features is less than that of low-frequency words.
5. High-frequency words are easier to understand than low-frequency words and their length is shorter, that's why high-frequency words in colloquial speech are produced as abbreviations.
6. High-frequency words are less loaded than low-frequency words from the point of view of information processing.
7. From the point of view of information theory, high-frequency words have less information than low-frequency words, because they have more predictability in text and speech.
8. High-frequency words are used more in standard language compared to low-frequency words, and low-frequency words are used in supra-standard and sub-standard language.

9. High-frequency words include grammatical words such as auxiliary verbs, prepositions, and conjunctions, common nouns and pronouns, while low frequency words include non-auxiliary verbs and those nouns that have a special meaning.

As for the Frequency Dictionaries, the first frequency dictionary was published by Kaeding (1897) in German language. He used a corpus of eleven million words to obtain the distribution of letters and letter strings in the German language. The first frequency dictionary for English was published by Ogden (1930 cited in Cermak and Kren, 2005) which was published with the aim of teaching English as a second language. Hasani (2005) using a text corpus of one million words prepared a frequency dictionary for the contemporary Persian language, which included 8000 lexical and non-lexical words.

Two key factors contribute to the widespread use of frequency dictionaries:

• theoretical factor: they offer fascinating perspectives on the language's lexicon, highlighting its core (the most often used terms) and its peripheral;

• Practical aspect: they may be applied immediately in real-world situations, particularly when choosing entries for brand-new bilingual or monolingual dictionaries.

As mentioned above as there are various psycholinguistic experiment in each of which frequency is a determining factor this article confines the analysis to the lexical decision experiments and the significant role of frequency effects in these experiments below. We can use a library of texts with a range of writers, styles, and genres. We could determine the true frequency of every word in the language if we could compile all of the written and spoken texts that now exist. Naturally, this is not feasible; instead, we must make due with a linguistic corpus, or sample of texts.

More trustworthy information about the language may be deduced from a larger corpus. However, there are other factors influencing the outcomes outside corpus size. It also relies on how the corpus is put together and how much of each component is there. For example, we would probably not acquire special terms—not even the most prevalent ones—if we limited the corpus to exclusively fiction. However, the quantity (frequency) of more frequent terms would become skewed if the only sources included were technical

reports or newspapers. Put otherwise, a representative corpus with a wide range of texts is necessary to encompass the majority of linguistic events, particularly lexical ones for our purposes.

3.2. Lexical Decision experiments

Lexical decision analysis is a technique for examining how quickly specific language items or structures are processed. Participants work at a computer running a specifically created program for processing speed estimation in this experiment. The participants hit a preset key on the keyboard to reveal their decision about word or non-word displayed on the screen, which is displayed by the computer one string of characters at a time. In order to determine the relative reading speed for each word and non-word, the program keeps track of the time spent pressing each key. Such an experiment's findings can be utilized to determine which aspects are more difficult to comprehend or take more time due to internal complexity or higher cognitive load (McEnery and Hardie, 2012).

3.2.1. Variables that correlate with lexical decision task:

3.2.1.1. Frequency

First, if a reliable frequency measure is applied, there is strong evidence that word frequency is the most significant variable in predicting lexical decision times and accuracy levels. Word frequency and response times (RT) had the strongest correlation across all databases. According to Yap and Balota (2009), it explained more than 40% of the variance in RT for mono-morphemic words in the English Lexicon Project (both monosyllabic and multisyllabic, see also Brysbaert & New, 2009). According to Ferrand et al. (2010), frequency might explain for 38% of the variance in RTs in the French Lexicon Project. For the Dutch Lexicon Project, Keuleers, Diependaele, and Brysbaert (2010) reported comparable numbers, while for the British Lexicon Project, Keuleers et al. (2012) reported similar numbers.

3.2.1.2. Prevalence

A new measure called word prevalence was recently presented by Keuleers, Stevens, Mandera, and Brysbaert (2015) and has the potential to explain a significant portion of the variation. They began with the finding that, based on their corpus frequency, some low frequency words appear to be more well-known than would be expected. Words that are widely known are more likely to be created often and, thus, to be encountered more frequently than words that

are known by a small number of people. This is the simplest way to understand the word prevalence effect, which has its origins with the word frequency effect. Some words were much better known to men than to women and other words were more prevalent among women. English equivalents of gender-specific words are *paladin, kevlar, dreadnought, and golem* (better known by men) and *tresses, taupe, peony, and bodice* (better known by women). This interpretation states that the word prevalence measure fills in the gaps in the corpus materials used to calculate the word frequency measurements. In fact, not all linguistic registers are represented in corpora (Brysbaert et al, 2015).

The word prevalence effect also appears to adjust for the familiarity with everyday things that many people know but which are rarely included in language corpora. There appears to be a significant amount of high-prevalence, low-frequency terms that are borrowed from other languages. Low frequency terms in science have a lot of English, French, or German roots. As a consequence, even if these terms are uncommon in a language, those who are familiar with these languages are probably going to comprehend them. Lastly, a large number of words with disparate word prevalence and frequency measurements are derived and complex words, which have low occurrence frequencies but are easily understood due to their decomposition. Words with high prevalence but low frequency (and familiarity) include distinctly, antioxidant, microbiology, antiviral, reusable, legalization, unsaturated, relenting, and preconditioned. Examining the degree to which morphological complex terms inherit the prevalence of their base words and the variables influencing the generalization will be an intriguing use of the word prevalence measurements (Brysbaert et al, 2015).

3.2.1.3. Word length

The quantity of letters and syllables in a word are two more factors that appear to be significant:

Words with several syllables and big words require longer word processing times.

3.2.1.4. Age of acquisition

A variable coming close to word frequency in the percentage of variance accounted for, is age-of-acquisition (AoA), the age at which the word was learned first. AoA is highly correlated with word frequency.

3.2.1.5. Similarity to other words

When making lexical decisions, words with similar orthographic patterns are processed more quickly than those with distinct letter sequences. This effect is often explained by postulating that the overall activity in the mental lexicon serves as a basis for some lexical judgments.

3.2.1.6. Concreteness

Although these effects are often considerable, semantic factors also predict a relatively tiny proportion of extra variance in lexical decision times. Juhasz, Yap, Dicke, Taylor, and Gullick (2011) found that less than 1% of the extra variation was explained by the semantic variables imageability, body-object interaction ratings, and sensory experience ratings (which indicate how much a word's meaning involves sensory experiences).

4. Data Gathering Regime and Corpus Compilation

4.1. Using Two Persian Frequency Dictionaries in Lexical Decision Tasks

Routledge Persian Frequency Dictionary (RPFD), which lists the 5,000 most commonly used terms in this language, is a good resource for anyone learning Persian. Based on a 150-million-word corpus of spoken and written Persian texts from Iran speech community, the dictionary offers the user part-of-speech and alphabetical indices in addition to a comprehensive frequency-based list. Each entry includes the translation into English as well as an illustration of usage in context. The Dictionary also includes lists of frequently used terms on a range of topics, arranged topically. In addition, lists of basic verbs and light verb formations as well as comparisons of various methods to represent the months of the year are included, with an emphasis on grammar. The dictionary offers a wealth of resources for creating curricula and teaching languages. The dictionary is a valuable tool for curriculum design and language instruction, and it is also available on a separate CD that offers the entire text in a tab-delimited format that is perfect for corpus and computational linguists. Students of all levels may effectively and entertainingly advance their study of Persian with the help of a Persian Frequency Dictionary.

It is the first Persian frequency dictionary, offering learners and other Persian language students a trustworthy core vocabulary based on a balanced sample of both spoken and written language.

"We sought to represent both text and speech because Persian exhibits a notably wider gulf (arguably diglossic) between the language traditionally used in writing ('standard') and that used in speech ('colloquial'), even among educated speakers, than one might encounter in the main varieties of English. As a kind of bridge between the text sources and the speech sources, we developed an online corpus of blogs. Online material can be referred to as Computer Mediated Communication (CMC), and has been found to represent a kind of middle ground between traditional text and speech (Warschauer, 2006)." (Miller and Aghajanian Stewart, 2018:1).

The dictionary indicates the raw frequency value which shows the number of total word family members associated with each headword that were counted in the corpus. Next comes the dispersion code whose value indicates whether the word is well-distributed throughout the corpus or not.

As for *frequency dictionary according to a written corpus of today Persian language (PFD),* a Persian corpus with the volume of 10618187 pieces of writing was used, which was distributed in 2990 text files. The texts of the body of corpus are selected from various sources such as books, newspapers, magazines, daily notes and others. The texts of the corpus of dictionary include written and spoken data, but their written or spoken characteristics are not given in the entries. The dictionary is organized alphabetically and not by frequency, and at the end, the words are arranged based on the frequency. The frequency mentioned for each entry is actually the sum of the frequency of the words under the entry. The words listed under each entry are either written differently or labeled differently, or both. For each word, first the frequency of the word in the dictionary body and then the tag of that word is mentioned. The words under the entry are sorted in descending order based on the frequency.

4. 1. 1. Prevalence taken into consideration

Word frequency is not a perfect estimate of word knowledge. For depicting this consider the Persian words listed below which have been chosen at random:

bozorg (big), besyār (many, much, very), ?āli (excellent), bad (bad), xub (good).

The frequency dictionaries -*Routledge Frequency dictionary of Persian* (Miller and Aghajanian-Stewart, 2018) and a *frequency*

*dictionary of Persian (*Bijankhan and Mohseni, 2018) - give us the frequency for each word as table (1) shows. The question that is raised here is whether the word frequency is a perfect estimate of processing time or not. Let us examine the hypothesis by a lexical decision test.

One of the methods of studying online language processing is using the multi-sensory decision making method which is famous by the terms *cross modal lexical decision (CMLD)* method firstly used by Shapiro and Levine (1990). Through this method, the processing of the sentence can be clearly shown, because the tasks in this method are sensitive to the moment-by-moment processing of the sentence. The research method here is multi-sensory lexical decision-making, in which the two senses of hearing and vision are simultaneously involved in the form of two tasks with the intention of understanding the meaning of sentences and lexical decision-making. The mentioned method has several features, which are:

1- The sentences that we are looking to investigate how they are processed are presented to the subjects in an auditory way. They are told that their main task is to listen to the sentences they hear.

2- The subjects are told that they have to do another task meanwhile, and that is, while they are listening to the sentence, they must decide about whether or not a chain of letters that appears on the screen of the computer is a word or a non-word. The logic of using the secondary task, which is mainly a lexical decision making, is that in this method, the time of doing the second task will reveal the processing time of the specific words in the heard sentence as there is a kind of trade-off between the two tasks (listening and deciding) in terms of their cognitive load across the mind.

3- There is no interruption in the processing of the sentence, even when the subject come across the visual stimulus (word/non-word).

4- Reaction time to visual stimuli reveals the processing time of the heard sentence. The method of data analysis in the test is based on the response time of the subjects to the visual stimuli. This work is calculated and recorded using response time measurement software based on thousandths of a second. This test is performed using the DMDX software program, which measures the reaction time of the subjects and their correct and incorrect answers in thousandths of a second. The results are calculated using appropriate statistical

methods. The results will be displayed in tables using Word and Excel programs.

This test is performed using the DMDX software program, which measures the reaction time of the subjects and their correct and incorrect answers in thousandths of a second. This program is mostly used by cognitive psychologists and language psychologists to analyze and evaluate different aspects of cognitive and language abilities.

table (1): Frequencies and RTs

	Number of subjects with correct answers	PFD Frequency	RPFD Frequency	Mean RT (ms)
bozorg (big)	20	409	9719	1193
besyār (many, much, very)	20	6380	14140	1070
?āli (excellent)	21	1382	3727	1061
bad (bad)	20	465	1845	1020.9256
xub (good)	21	1623	8010	1110.6449

As table (1) shows the same list of words carry different frequency estimates for the same specific list of words based on PFD and RPFD. As the numbers of table (1) show besyār enjoys the most frequent word among the five. But this word does not captures the fastest reaction time by the participants. This mismatch is highlighted when encountering some words in Persian which undergo the morphological process of conversion in which there is observed a change in the part of speech of words, without any change in the form of the word. Now the participants' reaction time during lexical decision can be seen[1]:

1. Doxtar-e bozorg* so? āl porsid. *girās - / adjective/ frequency: 409*
   daughter-EZ old- comparative marker question asked.
   The older doughter asked a question.

[1] . The sign of asterisk in the following test sentences show the exact location of presentation of visual stimuli. Next to each test sentence there have been inserted the word/non-word string of letters shown to the participants, the part of speech of the intended word whose speed of processing is to be calculated and its frequency in the corpus.

2.? in mas?aleh rā bozorg* jelveh dād. *dastur+/ adverb/ frequency:9*
This problem object marker bigger showed.
He made the issue appear big.

3. ānhā be bozorg*hā ehterām gozāštand . *be?temād -/ noun/ frequency: 2*
They to old- comparative marker respected.
They respected the older.

Table 2: descriptive statistics

	Number of subjects with correct answers	Mean reaction time	Standard deviation
1	20	1193.8702	621.90261
2	21	932.9260	680.10757
3	21	1145.7799	732.79054

The results of Friedman Test shows that there are significant differences between the mean sums at the significance level of 0.1:

Ranks

	Mean Rank
VAR00025	2.45
VAR00026	1.45
VAR00027	2.10

Test Statistics[a]

N	20
Chi-Square	10.300
df	2
Asymp. Sig.	.006

a. Friedman Test

We use Wilcoxon's non parametric test to compare questions. The results of the Wilcoxon test show that there is a significant difference between the mean of question 1 and the mean of question 2 at the significance level of 0.05. The word *bozorg* in the role of adjective is processed later than adverb and noun. This is the opposition to the presupposed assumption that the more frequent word will be processed sooner, too.

Test Statistics[a]

	VAR00011 - VAR00010	VAR00012 - VAR00010	VAR00012 - VAR00011
Z	-.403[b]	-1.198[c]	-2.016[c]
Asymp. Sig. (2-tailed)	.687	.231	.044

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

4.Tajrobe-(y)e besyār* xubi bud. *girās - / adverb/ frequency: 6380*
Experience- (hiatus) EZ very good was.
That was a very good experience.

5. ?in deraxt šāxe-hā-(y)e besyār* dārad . *tulāni+/ adjective/ frequency:640*
This tree branch- pl. many has.
This tree has many branches.

6. Barāy-e besyār*- i ?i:n mored piš mo? āyad. *e?temād -/ noun/ frequency: 143*
For- EZ many this case come about/ happen.
This case may happen to many.

Table 3: descriptive statistics

	Number of subjects with correct answers	Mean reaction time	Standard deviation
4	20	1070.5881	1075.57951
5	20	945.0402	832.55088
6	18	1076.5191	693.91432

The results of Friedman Test shows that there are significant differences between the mean sums at the significance level of 0.5:

Ranks

	Mean Rank
VAR00064	1.70
VAR00065	2.55
VAR00066	1.75

Test Statistics[a]

N	20
Chi-Square	9.100
df	2
Asymp. Sig.	.011

a. Friedman Test

The results of the Wilcoxon test show that there is a significant difference between the mean of question 1 and the mean of question 2 at the significance level of 0.05. The word *besyār* in the role of adverb is processed later than adjective and noun. This is the opposition to the presupposed assumption that the more frequent word will be processed sooner, too.

Test Statistics[a]

	VAR00074 - VAR00073	VAR00075 - VAR00073	VAR00075 - VAR00074
Z	-2.013[b]	-.450[c]	-1.586[c]
Asymp. Sig. (2-tailed)	.044	.653	.113

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

c. Based on negative ranks.

7. Modirān-e ?āli*-e sāzmān ?in tasmim rā gereftand. *girās - / adjective/ frequency: 1382*

Managers-EZ highest-EZ organization this decision made .

The highest managers of the organization made this decision.

8. ?āli* amal kard. *jāyeze+/ adverb/ frequency:4*

Excellent acted.

He acted excellently.

9. Az mobtadi tā ?āli* dars midād. *mošāhedeh + / noun / frequency: 2*

From beginner to advanced taught.

He taught from beginners to the advanced.

Table 4: descriptive statistics

	Number of subjects with correct answers	Mean reaction time	Standard deviation
7	21	1016.3771	506.26468
8	21	848.2170	397.00922
9	20	1001.73	932.642

The results of Friedman Test show that there are significant differences between the mean sums at the significance level of 0.5:

Ranks

	Mean Rank
VAR00070	2.45
VAR00071	1.75
VAR00072	1.80

Test Statistics[a]

N	20
Chi-Square	6.100
df	2
Asymp. Sig.	.047

a. Friedman Test

The results of the Wilcoxon test show that there is a significant difference between the mean of question 1 and the mean of question 2 at the significance level of 0.05. The word *āli* in the role of adverb is processed sooner than adjective and noun. This is the opposition to the presupposed assumption that the more frequent word will be processed sooner, too.

Test Statistics[a]

	VAR00071 - VAR00070	VAR00072 - VAR00070	VAR00072 - VAR00071
Z	-1.686[b]	-1.568[b]	-.261[b]
Asymp. Sig. (2-tailed)	.092	.117	.794

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

10. Hameh bad* nistand. *ʒangin-/ adjective/ frequency: 465*
All bad* aren't.
It is not the case that all (of them) is bad.

11. Be u: bad* hamleh kardand. *tozi?+/ adverb/ frequency:37*
To him bad attack did.
They attaked him badly.

12. Az bad*-e hādese injā?-im tamite *-/ noun/ frequency: 10*
from the state of being bad-EZ the event here- personal ending for 1sing.
We are here as a result of a misfortune.

Table 5: descriptive statistics

	Number of subjects with correct answers	**Mean reaction time**	**Standard deviation**
10	21	1020.9256	498.26265
11	18	1077.1885	1179.94380
12	21	1392.9256	1053.37167

The results of Friedman Test show that there are significant differences between the mean sums at the significance level of 0.05:

Ranks

	Mean Rank
VAR00022	2.06
VAR00023	1.50
VAR00024	2.44

Test Statistics[a]

N	18
Chi-Square	8.111
df	2
Asymp. Sig.	.017

a. Friedman Test

We use Wilcoxon's non-parametric test to compare questions. The results of the Wilcoxon test show that there is a significant difference between the average of question 3 and the average of questions 2 and 1 at a significance level of 0.05. That is, the word *bad* is processed later as a noun than as an adjective or adverb.

Test Statistics[a]

	VAR00023 - VAR00022	VAR00024 - VAR00022	VAR00024 - VAR00023
Z	-.414[b]	-2.555[c]	-1.851[c]
Asymp. Sig. (2-tailed)	.679	.011	.064

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

c. Based on negative ranks.

13. šārāyet-e xub* va monāsebi-I dāšt. *?adabiyat +/ adjective/ frequency:1623*
Conditions-EZ good and suitable-indefinite art. had (he had).
They determined a stable and permanent location.

14. Bezan ke xub* mizan-i. *faqsim - / adverb / frequency:674*
Play for well play-personal ending 2sing.
Play because you play (the piano) well.

15. In xub*-hā va bad-hā dar mo?qe?iyat-hā bā ham farq dārand. *ma?ruf +/noun/ frequency: 20*
This good-pl. and bad-pl. in situation-pl. with each other are different.
These good and bad things are different in different situations.

Table 6: descriptive statistics

	Number of subjects with correct answers	Mean reaction time	Standard deviation
13	20	1110.6449	1065.84928
14	21	955.8641	397.22910
15	21	1087.6488	1157.32366

The results of Friedman Test shows that there are significant differences between the mean sums at the significance level of 0.05:

Ranks

	Mean Rank
VAR00064	1.70
VAR00065	2.55
VAR00066	1.75

Test Statistics[a]

N	20
Chi-Square	9.100
df	2
Asymp. Sig.	.011

a. Friedman Test

We use Wilcoxon's non-parametric test to compare questions. The results of the Wilcoxon test show that there is a significant difference between the mean of question 2 and the mean of question 3 at a significance level of 0.05. That is, *xub* is processed as a noun later than the adjective. Although *xub* as adjective is much more frequent than *xub* as adverb and noun, it is processed later than both.

Test Statistics[a]

	VAR00065 - VAR00064	VAR00066 - VAR00064	VAR00066 - VAR00065
Z	-.896[b]	-.224[b]	-2.312[c]
Asymp. Sig. (2-tailed)	.370	.823	.021

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

The analysis of the informative content of the above tables show that frequency may not be considered as an ultimate sole estimate measure for processing time, for other factors may affect it. Now these factors will be reviewed:

4.1. 2. The interaction of frequency with other processing factors such as recency and salience

It is assumed that frequency interacts with other processing factors such as recency or salience.

Recency: The timing of frequency effects determines their impact. The processing of a present speech event is more influenced by a speech event that occurred recently than by one that occurred earlier. This impact can be directly derived from memory structure and has been observed in language processing on several occasions (Szmrecsanyi 2006; Ellis 2012c). This is especially important when there are two free variations of a structure or category in a given situation.

Three things matter in this situation: First, the time passed since the item last happened: the greater the activation of an item, the more recently it has occurred (Szmrecsanyi 2006). Second, the number of occurrences of an item during a given period of time: the greater the frequency of recent occurrences of an item, the stronger its activation. Thirdly, the item's overall frequency: an item's recency impacts are larger the lower its general token frequency (Jacoby and Dallas 1981; Schwenter 2013). Low token frequency increases the recency effects since it is associated with a higher rate of surprise, which strengthens activation.

The time of the exposure also affects recency effects. When it comes to memory retention, spaced learning has a general benefit over massed learning. A more dispersed exposure where the experience is reactivated appears to be more beneficial than being exposed to a large number of tokens in a little amount of time.

Salience: According to Tomlin and Myachykov (2015), frequency effects are also impacted by the salience of the corresponding grammatical or phonetic unit. Prosodic salience is one definition of salience. Speech events that are more prosodically pronounced—that is, more rhythmical—activate words more powerfully than ones with less pronounced prosody. The main fields of study for morpho-

syntactic salience include sociolinguistics, perceptual dialectology, and dialect contact.

4.2.3. Type and Token frequency ignored

Two main types of frequency effects have been described in the literature: token frequency and type frequency. We distinguish explicitly between type and token frequency impacts here. Token repetition or frequency on its own results in entrenchment rather than generalization (cf. Bybee 2006, 2010; Ambridge et al. 2015). However, in order to serve as the foundation for potential schema construction and generalization, type frequency or variation is required (Langacker 1987; Bybee 2006, 2010; Ambridge et al. 2015). Reinforcement of memory traces results from repeated encounters (see Blumenthal-Dramé (2012), Divjak and Caldwell-Harris (2015). Several intriguing processes result from the entrenchment of linguistic units. Since high frequency units are usually easier to get, this improves representation stability and makes retrieval easier. Two main types of frequency effects have been described in the literature: token frequency and type frequency. Each of these gives rise to the entrenchment of different kinds of linguistic units. While token frequency gives rise to the entrenchment of instances, type frequency gives rise to the entrenchment of more abstract schemas. Token frequency refers to the frequency with which specific instances are used in language.

Regarding the idea of frequency, the most fundamental contrast is between token frequency and type frequency (cf. Ellis 2012c). The number of times a concrete form (or a lemma) appears in a corpus or in the input overall is referred to as token frequency. For instance, how frequently does the word form played appear in a corpus? Elevated token frequency usually results in the entrenchment of a specific instantiation of the structure rather than high productivity of the construction itself.

On the other hand, type frequency describes the quantity of unique things that can occupy a slot in a specific construction: For instance, how many distinct vocabulary items may be used with the English Past Tense Construction VERBed? Determining a construction's type frequency is essential to figuring out how productive it is.

5. Data Analysis and Discussion

Here, the main questions were: How can the methods and findings of corpus linguistics and psycholinguistics continue to usefully

interact in the course of frequency dictionaries? How can corpus methods and findings be utilized in the analysis of the textually mediated world in psycholinguistics?

In many psycholinguistic researches there is a need for (auditory or visual) stimuli which are matched across frequency of occurrence in order to minimize and control the frequency effect which can interfere with the other specified factors and variables. That is why frequency dictionaries are of significance and use in psycholinguistic researches, especially in researches which are mainly based on lexical decision tasks. In Persian there are two main frequency dictionaries (called here, RPFD and PFD) which have been used in some psycholinguistic researches by the authors of this article (Mehrabi, 2014; 2015; Mehrabi and Zaker, 2013; Mehrabi and Mahmoodi-Bakhtiari,2020, 2021a, 2021b, 2022a,2022b; Mehrabi et al. 2021) which are mainly based on lexical decision tasks. Using this kind of frequency dictionaries they have confronted some difficulties which can be listed as follows:

1. As the sources determined for gathering data in order to compile a corpus are different in these two dictionaries different frequency estimates are presented from the same list of words. Here, the examples for this claim were given through five randomly chosen words of *bozorg* (big), *besyār* (many, much, very), *?āli* (excellent), *bad* (bad), *xub* (good). Although these two frequency dictionaries have been published in 2018, they present different frequency estimates summarized in table (1). The only similarity is that *besyār* in both dictionaries ranks the much more frequent word than the other four words.

2. Through the process of conducting the researches it was found that word frequency may not be a perfect estimate of word knowledge and processing time. For examining this hypothesis an experiment was conducted in which the relation of processing time and frequency of words have been evaluated. Here, the direct relation of frequency and processing time was re-examined by a lexical decision task done on these five specific words whose results can be witnessed in tables 2-6. This direct relation was observed only in the case of lexical decisions related to words *bad* (bad), *xub* (good). This mismatch between the frequency and processing time shows that the researcher of psycholinguistics must take other intervening psychological factors such as recency, salience, and

word prevalence into consideration as he is choosing word lists for conducting lexical decision tasks.

3. The presupposed assumption of the existence of the direct relation between frequency and processing time may be opposed in Persian especially when confronted to some words which can undergo the lexical process of conversion (when new words are formed simply by changing their grammatical function in the speaker's mind while the external form of the word does not undergo any special phonetic change and without receiving an affix, it undergoes derivation and is used for different syntactic contexts). Here, the five chosen Persian words may be converted. The results of the analysis of above tables show that the opposition to the presupposed assumption that the more frequent word will be processed sooner can be conveyed. For instance, as table (2) shows the word *bozorg* although more frequent as an adjective, is processed later than the same word as an adverb or a noun.

6. Concluding Remarks

Corpus-based psycholinguistic tests will not achieve their ultimate goal of revealing the mental processes unless corpora have been compiled with noticing some psychological findings. Here, it was shown that for the existence of intervening cognitive/ processing factors like prevalence, salience, and prominence frequency –yet first factor affecting ease, speed, and accuracy of linguistic item processing- is not the sole factor that must be considered while examining word knowledge of the participants. Taking into consideration of those factors may explain why in some cases word frequency is a sufficient estimate for evaluating word knowledge through lexical decision tasks. It will enhance the content of frequency dictionaries with more psycholinguistic real information and increase their validity by adding new lists that may show word prevalence and type frequencies.

It seems that frequency dictionaries are formed and framed solely based on the word encounter in the corpus, while there is a need for taking into consideration of some cognitively real and psychologically significant factors like word prevalence, word salience (prosodic prominence), or recency effect. The recency effect can be considered in this way; frequency dictionaries should be updated regularly (e.g., annually).

Frequency dictionaries of Persian don't consider this significant issue that there were also interesting differences between genders. Some words were much better known to men than to women and other words were more prevalent among women or some words are best known in academic jargon than the speech of the ordinary people. This may affect frequency counts.

As the source of corpus compilation in these dictionaries are mainly based on written materials a big portion of speech mainly spoken language is ignored. The corpus must represent a balance between text and speech and cover a variety of genres. Type frequencies seem to be ignored in these dictionaries specially in RPFD.

6.1. Ways to Overcome the Challenges

To resolve these problems the psycholinguistic researcher must

a. consider the type of resources used as the basis of frequency dictionary which is going to be used, for the resources affect frequency counts.

b. conduct a word knowledge test to the participants before choosing the lexical decision stimuli. This word knowledge test will depict which words the participants know much better than the other words. This will give the researcher an insight of the participants' word knowledge and will help the researcher to choose much more suitable words for lexical decision task.

c. enjoy updates. Frequency dictionaries of Persian don't enjoy soon updates. These frequent updates are necessary for meeting recency effect.

d. include prevalence information. Persian frequency dictionaries can include prevalence information in the form of appendices at the end of their word lists.

To summarize empirical results, the analysis of the informative content of the above tables show that frequency as depicted and measured in the two above mentioned Persian frequency dictionaries may not be considered as an ultimate sole estimate measure for processing time, for other factors may affect it. To review these intervening factors one must consider prevalence, recency, and salience. Other concise researches need to be done in order to investigate how and to what extent this goal is achievable. Here is where psycholinguistics must contribute to the corpus linguistics.

References

– Ambridge, B., Kidd, E., Rowland C. F, & Theakston, A.L. (2015). The Ubiquity of Frequency Effects in First Language Acquisition. *Journal of Child Language* 42(2): 239–273.

– Bijankhan, M. & Mohseni, M. (2018). *Frequency Dictionary According to a Written Corpus of Today Persian Language*. Tehran: University of Tehran.

– Blumenthal-Dramé, A. (2012). *Entrenchment in Usage-Based Theories: What Corpus Data Do and Do Not Reveal About The Mind*. Berlin/New York: de Gruyter.

– Brysbaert, M., & Ellis, A. W. (2016). Aphasia and Age of Acquisition: Are Early-learned Words More Resilient? *Aphasiology, 30,* 1240–1263.

– Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods, 41,* 977–990.

– Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review: *Current Directions in Psychological Science*. vol. 27 (1).

– Brysbaert, M., Mandera, P., & Stevense, M. (2015). The Impact of Word Prevalence on Lexical Decision Times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology Human Learning & Memory.*

– Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language,* 82(4): 711–733.

– Bybee, J. L. (2010). *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.

– Cermak, F. & Kren, M. (2005). New Generation Corpus-based Frequency Dictionaries. *International Journal of Corpus Linguistics*, 4. Pp. 453-467.

– Divjak, Dagmar & Catherine L. Caldwell-Harris. (2015) Frequency and entrenchment. In: Ewa Dabrowska and Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 53–75. Berlin/ New York: de Gruyter.

– Ellis, N. C. (2012). What Can We Count in Language, and What Counts in Language Acquisition Cognition, and Use? In: Stefan Th. Gries and Dagmar S. Divjak (eds.), *Frequency Effects.*

– Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Meot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: Lexical Decision Data for 38,840 French Words and 38,840 Pseudowords. *Behavior Research Methods, 42,* 488-496.

– Gimenes, M., and New, B. (2016). Worldlex: Twitter and Blog Word Frequencies for 66 Languages. *Behavior Research Methods*, *48*, 963–972.
– Herdag˘delen, A., & Marelli, M. (2017). Social Media and Language Processing: How Facebook and Twitter Provide the Best Frequency Estimates for Studying Word Recognition. *Cognitive Science, 41,* 976–995. doi:10.1111/cogs.12392.
– Hasani, H. (2005). *The Most Frequent Contemporary Persian Words*. Tehran: Iran Language Center.
– Jacoby, L. L. & Dallas, M. (1981). On the Relationship between Autobiographical and Perceptual Learning. *Journal of Experimental Psychology: General* 110: 306–340.
– Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2016). Experience as a Free Parameter in the Cognitive Modeling of Language. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2291– 2296). Austin, TX: Cognitive Science Society.
– Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible Words Are Recognized Faster: The Grounding of Meaning in Sensory and Perceptual Systems. *The Quarterly Journal of Experimental Psychology, 64(9),* 1683-1691.
– Kaeding, F.W. (1897). Haufigkeitsworterbuch der DeutschenSpreche. SteglitzEigenverlag.
– Keuleers, E., Diependaele, K. & Brysbaert, M. (2010). Practice Effects in Large-scale Visual Word Recognition Studies: A Lexical Decision Study on 14,000 Dutch Mono- and Disyllabic Words and Nonwords. *Frontiers in Psychology 1:174*. doi: 10.3389/fpsyg.2010.00174.
– Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical Decision Data for 28,730 Monosyllabic and Disyllabic English Words. *Behavior Research Methods, 44*, 287-304.
– Keuleers, M., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word Knowledge in the Crowd: Measuring Vocabulary Size and Word Prevalence in a Massive Online Experiment. *Quarterly Journal of Experimental Psychology*.
– Langacker, R. W. (1987). *Foundations of Cognitive Grammar*, Vol. 1: *Theoretical Prerequisites.*Stanford: Stanford University Press.
– Miller, C. & Aghajanian-Stewart, K. (2018). A *Frequency Dictionary of Persian, Core Vocabulary for Learners.* London: Routledge.
– McEnery, T., & Hardie, A. (2012). *Corpus Linguistics; Method, Theory, and Practice*. Cambridge: Cambridge University Press.

– Mehrabi, M. (2014). Lexical Information of Persian Transitive Verbs during Sentence Comprehension. *Language Related Research.* Vol. 5, No. 1, pp. 271-295.
– Mehrabi, M. (2015). The Effect of Farsi Verb Representational Complexity on Processing Time During Listening Comprehension. *Journal of Researches in Linguistics*, Vol. 7(1), Pp.77-92.
– Mehrabi, M. & Mahmoodi-Bakhtiari, B. (2020). A Comparative Study of Comprehension of Evidentiality in Persian, English, and Turkish; a Psycholinguistic Approach. *Comparative Linguistic Researches.* N. 20. Pp. 243-257.
– Mehrabi, M. & Mahmoodi-Bakhtiari, B. (2021a). The Psychological Reality of Evidentiality Hierarchy in Persian during Sentence Listening Comprehension. *Language Related Research*. N.12. Pp. 539-566. https://doi.org/10.29252/LRR.12.2.17
– Mehrabi, M. & Mahmoodi-Bakhtiari, B. (2021b). Mental Representations of Persian and English Absolute and Relative Tenses: A Contrastive Psycholinguistic Approach. *Journal of Researches in Linguistics*, Vol. 13(1), Pp.89-112.
– Mehrabi, M. & Mahmoodi-Bakhtiari, B. (2022a). Representational Complexity of Persian Absolute Tenses during Listening Comprehension. *Scientific Journal of Language Research*, Vol. 13, No. 41. Pp.55-80.
– Mehrabi, M. & Mahmoodi-Bakhtiari, B. (2022b). Representational Complexity of Persian Relative tenses during listening Comprehension. *Language Related Research.* Vol. 13, No. 2, Pp. 1-32. https://doi.org/10.52547/LRR.13.2.1
– Mehrabi, M. & Mahmoodi-Bakhtiari, B. & Vaezi, H. (2021). Auditory Perception of Persian Interrogatives from a Psycholinguistic Approach and its Application in Persian teaching. *Journal of Teaching Persian to Speakers of Other Languages*. Vol. 10, No. 2. Pp. 137-156.
– Mehrabi, M. & Zaker, A. (2013). Lexical Information of Persian Transitive Verbs during Sentence Comprehension. *Iranian Studies.* Vol. 46, No. 6, Pp. 959–971.
– Miller, C., & Aghajanian-Stewart, K. (2018). *A Frequency Dictionary of Persian; Core Vocabulary for Learners.* London: Routledge.
– Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of Frequency on Visual Word Recognition Tasks: Where Are They? *Journal of Experimental Psychology: General*, *118*, 43–71.
– Niemikorpi, A. (1997). Equilibrium of Words in the Finnish Frequency Dictionary. *Journal of Quantitative Linguistics*. Vol. 4. NO: 1-3. Pp. 190-196.

– Schwenter, S. A. (2013). *Strength of Priming and the Maintenance of Variation in the Spanish Past Subjunctive*. Paper Presented at NWAV 2012 Pittsburgh.

– Shapiro, L.P. & B. Levine. (1990). Verb Processing during Sentence Comprehension in Aphasia. *Brain and Language*, 38: 21-47.

– Szmrecsanyi, B. (2006). *Morphosyntactic Persistence in Spoken English: A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin/New York: de Gruyter.

– Tomlin, R. S. & Myachykov, A. (2015). Attention and Salience. In: Ewa Dabrowska and Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 31–52. Berlin/New York: de Gruyter.

– Warschauer, M. (2006). *Laptops and Literacy: Learning in the Wireless Classroom*. New York: Teachers College Press.

– Yap, M. J., & Balota, D. A. (2009). Visual Word Recognition of Multisyllabic Words. *Journal of Memory & Language, 60*, 502-529.

– Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*, 441–517.

– Zolfaghari, H. & Hasani Kochaki, E. (2023). Stylistic Analysis of Prose of Qajar Period Newspapers. *The Journal of Linguistic and Rhetorical studies.* Volume 14, Consecutive Number 31. Pp. 65-98.

About the Authors:

Masoumeh Mehrabi is a graduate in Linguistics and a faculty member at Ayatollah Ozma Borujerdi University. Her main areas of teaching and research include psycholinguistics, (critical) discourse analysis, the linguistic study of Persian literature, and the Persian language.

Behrooz Mahmoudi-Bakhtiari is a graduate in Linguistics and a faculty member at University of Tehran. His main areas of teaching and research include dramatic discourse analysis, theater and film semiotics, Persian language education, and Iranian dialects.