

*Full Length Research Paper*

# Students' preferences for syntax usage in Turkish language using distributional linguistic theory

Cem Erdem

Adam Mickiewicz University, Poland.

Received 11 October, 2016; Accepted 5 December, 2016

Communication is one of the most important aspects of social life. Social interactions have increased the necessity for communication and learning of language. Social needs which constitute the main goal of teaching activities fill the gap of learning language. Linguistic surveys have revealed an important finding on educational activities and outputs. One of them is distributional linguistic theory. The purpose of this work is to describe the connection between Polish speakers who learned Turkish as a foreign language having preference for syntax and Turkish native speakers having preference for syntax using distributional linguistic theory. In this study, survey model, a kind of quantitative research, was used. Quantitative research data were collected by document review. The survey was carried out with Polish and Turkish university students in 2013/2014 academic year in spring semester. The working group consists of 20 students who were chosen with stratified random sampling. The findings explain the meaning statistically, and consequences were tabulated. There are differences in the syntax and sentence length of the working group in written language.

**Key words:** Turkish language education, distributional linguistic theory, writing skill.

## INTRODUCTION

Meaning represents a concept which is greater than its components since it occurs in advanced and complex expressions. Meaning transformation activity which is the main purpose of communication takes place with words. Being the main components of sentences, words are the main elements which enable meanings to take place in the determined direction. Furthermore, meaningful features, obtained with words in a sentence, have determinant roles in shaping and diversifying meaning levels. Usages, which widen meaning limits of words during communication, and new meanings, which are obtained through syntagmatic roles within sentence, consist of the important part of the study area of

linguistics.

The studies conducted abroad about syntax gained new levels with advancement of linguistic field. Two schools draw attention. First is American structural linguistic school and second is productive transformational linguistic school (Işkın, 1999). Among these schools, Bloomfield analyzed the grammar of language in details and conducted important studies on the subject of syntax analyses. Therefore, he became a muse for structural science people, which came after him, using his approach of linguistic analyses which was called "direct components analyses".

These studies of Bloomfield enabled the establishment

E-mail: [cemerdem@gmail.com](mailto:cemerdem@gmail.com).

Authors agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

of “distributional linguistic” by Zellig Harris, who was not structural with subject of form analyses. They enabled the occurrence of productive transformational grammar which was developed by Chomsky (Demirel, 2011). Distribution approach, which is shaped in this foundation, found meaning with the thought that syntax and meaning are distributed in a similar fashion (Rubenstein and Goodenough, 1965). Indeed, with this approach, it is assumed that structure could be understood as a whole with distributional representations of word meanings throughout a sentence (Landauer and Dumais, 1997).

On the other hand, Chomsky intended to separate the sentences of a language, which are appropriate to grammar from the inappropriate ones and to reveal the structures of sentences, which are appropriate to its grammar, in linguistic analyses related to surface structure. Chomsky stated that there were mathematical formulas in terms of structure in sentence setting for all languages. The main purpose of linguistic analyses is to study which sentences are linguistically correct and which are not appropriate linguistically (Chomsky, 1965; Demirel, 2011, p.34).

Therefore, the findings obtained through syntax analyses enable one to define basic language data about the establishment of meaning units and ordering (Sahlgren, 2008). Likewise, Pantel (2005) stated that word syntax are prone to gain similar meanings in syntax. Firth (1957) added that it is important to know what word groups follow in the whole sentence. Generally, distributional structure argues that there is a close relationship between structure and meaning within a text (Sahlgren, 2008).

Saussure (1985) argued that parole is individual and langue is a societal concept. Therefore, langue should be analyzed. Saussure (1985) regards langue as a system in which the components, that create the langue, are in interaction with each other within specific rules (Demirci, 2009). According to Saussure (1985), paradigm is a plane on which we can make horizontal changes. On the other hand, syntagma is a plane on which we can make horizontal and vertical changes

On the other hand, Zellig (1968) focused on the concept of verbal language (Friedman, Kra, and Rzhetsky, 2002). He considered words with their different distributions on usage. Distributional approach means evaluation of independent components which created the meaning (Sahlgren, 2008). Through this context, it could be stated that distributional approach depends on the assumption that there is distribution and meaning similarities within the nature of language. After Zellig (1970) developed his studies, he understood the concept of distribution required of him to work infinite number of sentences than when he focused on the concept of “transformation”. Therefore, he tried to specify kernel sentences, since once kernel sentences could be found, the structure of language would be understood

with those, according to Zellig (1970), stated that this method could be applied to all languages (Vardar, 1999).

Four different bits of information could be obtained about language based situations with distributional analyses. They are as follows: potential usage of syntax, communicational or situational content, phonological idea in syntax classifications and prosody (Chater and Finch, 1998).

According to Zellig (1970), syntax based and word based distribution give general information about the whole text. The whole text could be understood with the distribution of concept (Piits, 2013). Redington et al. (1998) stated three syntax classifications in distributional approach:

1. Determining distribution of each word within the text
2. Comparing the distribution of word combinations in the text
3. Determining the words, which are similar and used together, in the text

Leonard Bloomfield stated that the structures, which we ignore since we do not know their meanings, could be understood through their distribution in the following sections of the text (Harris, 1970). Therefore, he claimed that the word groups, which are source of establishment of text components in linguistic manners, create the linguistic meaning (Harris, 1968).

Learning a foreign language is acquisition of new linguistic system, which is composed of meanings and sounds; to learn speaking rules related to this system and different types of communication functions; and use them in an appropriate and correct ways. In such a learning process, one could only be possible with a language teaching approach, which focuses on a person to comprehend the language cognitively and him/her to develop communicational competence (Matthews, 1997). Learners learn new system and its rules as utilizing language systems and their native language in order to develop language skills and they focus on competencies (Tura, 1983).

According to Redington et al. (1998), syntax distributions of words in sections of the text are defined as important sources in language acquisition and in defining the text (Redington et al., 1998). On the other hand, Turney and Pantel (2010) stated that specifying the syntax usage of individuals in statistic manner give answer to the question of which is understood. Again, in the biomedical studies, it is found out that distributional approach is effective in defining the natural usage of language. Then it is stated that the findings could be used with neuro-linguistic method (Weeds et al., 2005).

Not only the theories in the area (Sentaks) should be given, also the results of the research on the syntax should be included. In particular, the results of the studies emphasizing the relationship of syntax to writing skills

**Table 1.** Vertical and horizontal plane in sentence.

Syntagma		Paradigm	
My sister	Tomorrow	To London	(she will) Fly
Ali	Today	At us	(he will) stay
S/he	In the next week	At Ankara	-will be-
I	Tuesday	-From- Germany	(I will) leave

should be included (Miller G, Walter C (1991). In this study, revealing the relationship between syntax preferences of students, whose native language is Polish, and regarding Turkish as a foreign language, and syntax preferences of students whose native language is Turkish aims to make syntax preferences of students visible (Table 1).

## METHODOLOGY

Information on research model, study group and data collection process is provided in this section.

### Research purpose

The purpose of the study is to reveal the relationship between syntax preferences of students, whose native language is Polish, and regard Turkish as a foreign language, and syntax preferences of students whose native language is Turkish.

### Research model

In this study, survey model which is a kind of quantitative research was used. This method intends to analyze research questions with different perspectives, and to find clearer answers for them as using both quantitative and qualitative data. The data in quantitative section are collected through document analysis (Yildirim and Şimşek, 2008).

### Study group

The study group is composed of 20 students, who were selected with stratified random sampling method, and from Turkish Teaching Department of Gazi University in Turkey and Jagiellonian University, Warsaw University, and Adam Mickiewicz University in Poland, in spring term in 2013/2014 education year.

### Data analysis

Content analysis method is used for data analysis in this research which is conducted through document analysis method. Content analysis is a systematic and repeatable method in which some words of the text are summarized with smaller content categories, through coding depending on specified rules (Büyüköztürk et al., 2008). For this purpose, essays obtained from the study group are analyzed according to distributional approach analysis method. For quantitative analysis, dependent-independent groups t-test is used.

The students were asked to write a thought article. With a view to make it possible for students to express themselves in a better way no particular title was identified. It was explained that these articles did not have a purpose of evaluation as an exam and would not transform into exam marks but were made only to observe the situation; the students were asked to reflect the knowledge and skills regarding writing composition which they had been obtaining so far in the articles. This process was completed in 50 to 60 min.

## RESULTS

### Findings related to syntax

Item orders and sentence structures obtained from the students in Poland Turcology departments are analyzed as follows:

Findings in Table 2 show that Poland students set sentences, which are created by at most 5 items. Additionally, it is found out that amount of sentences which are composed on single and 5 items are preferred at the smallest rate, in similar ways.

It is seen that Poland students use average of  $\bar{X} = 9.18$  while Turkish students use average of  $\bar{X} = 10.20$  words and also Poland students tend to write essays with average of  $\bar{X} = 15.7$ ; while Turkish students with average of  $\bar{X} = 18.2$  sentences (Tables 3 and 4).

Findings show that sentence structures, which are greater than two item- sentences, used relatively low amounts in students' written communication. Findings show that Polish students prefer using sentences which are created with at most 2 items for Turkish sentences.

Findings show that the distribution has common features up to sentences which are created with three items. Additionally, Polish students use sentences with 4 items relatively less than Turkish students. Demirel (2011) shares similar information regarding that writing focuses on thought and students can realize thoughts.

From the necessary assumptions to make ANOVA test come true, whether distribution shows normality or not was examined by "Kolmogorov-Smirnov test" (Table 5). Being between -1 and +1 of these values verifies the hypothesis that study group marks show normal distribution (Büyüköztürk, 2013; Cohen et al., 2000). Towards the findings (Table 6) that were obtained, using

**Table 2.** Table for distribution and numbers of students within universities.

	Universities	Gender		Language levels	Total
		Female	Male		
Group	Jagiellonian University	2	1-	B2	3
	Warsaw University	1	2	B2	3
	Adam Mickiewicz University	2	2	B2	4
	Gazi University	6	4	C2	10
	Total	11	9		20
		55%	45%		100%

**Table 3.** Table for distribution and frequency of items within sentences.

Frequency of items within sentences	Sentence numbers (Poland)	Sentence numbers (Turkish)
<b>One item- sentences</b>		
V	4	8
<b>Two item- sentences</b>		
S+V	21	28
O+V	15	17
A+V	15	15
IO+V	9	10
<b>Three item- sentences</b>		
S+A+V	11	7
A+O+V	10	3
S+IO+V	6	9
A+S+V	5	3
S+O+V	4	23
IO+A+V	4	1
A+S+V	3	3
O+A+V	2	1
A+A+V	2	3
A+IO+V	2	1
IO+O+V	2	2
IO+S+V	1	4
O+S+V	1	1
<b>Four item- sentences</b>		
A+N+A+V	4	3
S+A+A+V	2	2
A+IO+A+V	2	1
A+A+IO+V	2	1
S+A+IO+V	3	1
A+S+A+V	3	1
A+IO+S+V	2	2
A+A+O+V	1	2
IO+S+IO+V	1	1

**Table 3.** Cont'd.

A+S+IO+V	1	1
A+A+S+V	1	1
S+IO+O+V	1	2
S+A+A+V	1	1
IO+S+A+V	1	1
IO+A+IO+V	1	-
A+IO+O+V	1	1
S+IO+A+V	1	5
<b>Five item- sentences</b>		
A+A+O+A+V	2	5
S+O+A+IO+V	1	4
S+A+O+A+V	1	1

**Table 4.** Table for Polish and Turkish students' word and sentence length.

	Average of word length	Average of sentence length
Polish students	9.18	10.2
Turkish students	15.7	18.2
Mean	12.44	14.2

**Table 5.** Tests of Kolmogorov-Smirnov and Shapiro-Wilk.

Variable Type	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Subject	0.12	20	0.21	0.95	20	0.38
Indirect object	0.11	20	0.23	0.94	20	0.26
Object	0.19	20	0.04	0.83	20	0.03
Adverbial	0.11	20	0.20	0.95	20	0.50
Verb	0.15	20	0.20	0.90	20	0.05
Average	0.18	20	0.07	0.93	20	0.21

parametric tests in the analysis of data obtained from study group was approved and the execution was practiced accordingly ( $p>0.05$ ).

Tables 6 and 7 shows that in the sentences, which are used by Polish students in essays, meaningful differences could not be found in terms of word averages per sentence ( $p>0.05$ ).

### Findings related to item usage

In this section, part of the first 1400 words is considered in texts in order to eliminate the calculation errors which

may be caused by the length of texts in the evaluations related to item amounts. In this purpose, a standard length is sustained in written communication of students. In the essays, which are created by Polish students in essays, meaningful differences could not be found in terms of item types ( $p>0.05$ ).

### Conclusion

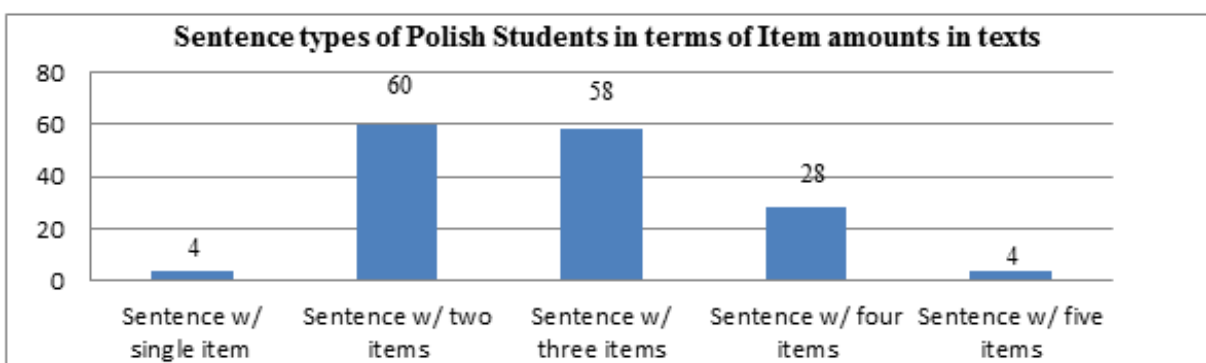
It is found out that, S-V, O-V and A-V ordering in sentences with two items; S-A-V ordering in sentences with three items; A-O-A-V, S-A-A-V ordering in sentences

**Table 6.** Independent sample T test results for word average variable per sentence.

Country	N	F	S	Sd	T	P
Turkey	10	1.939	0.181	19	1.374	0.186
Poland	10					

**Table 7.** Independent T Test Results Related to Item Variable

Variable Type	F	S	Sd	T	P
Subject	0.167	0.689	19	0.761	0.45
Indirect Object	1.306	0.271	19	0.365	0.72
Object	1.134	0.432	19	0.786	0.36
Adverbial	1.754	0.205	19	1.159	0.26
Verb	0.337	0.570	19	1.869	0.80

**Graph 1.** The amount of sentences used by Polish students in written communication.

with four items; and A-A-O-V ordering in sentences with five items are used frequently in sentence structures in essays of Polish students. On the other hand, it is found out that S-V in sentences with two items; S-O-V ordering in sentences with three items; S-IO-O-V ordering in sentences with four items are used frequently in sentence structures in essays of Turkish students. In this context, differences are established in sentence structures of sentences, which are created with three and four items.

It was found out Polish students use sentences, which are created with 2 and 3 items, relatively more frequently in written communication. It is found out that Polish students prefer using sentences, which are created with 4 items, less as compared to Turkish students. It could not find a meaningful difference in essays belonging to Poland and Turkish students in terms of item type frequency.

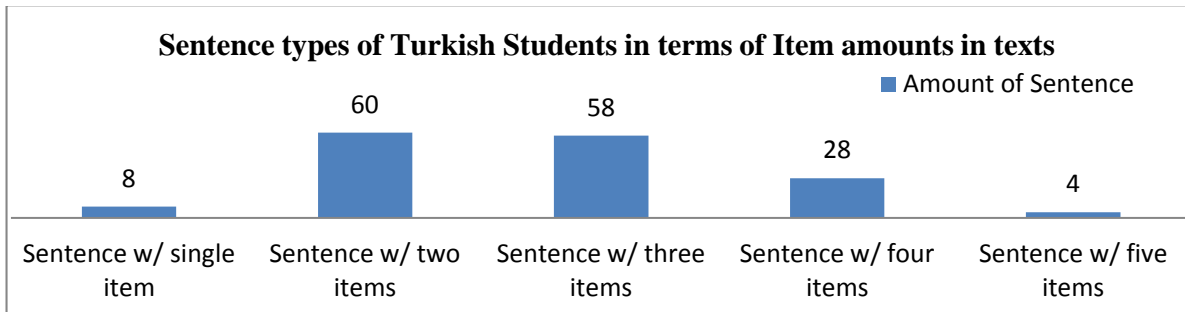
Adverbial item is used more frequently comparing to other items in sentences of Polish students.

It is seen that the sentences of Poland students have similarities with Turkish students' sentences in terms of distribution of item amounts (Graphs 1, 2 and 3).

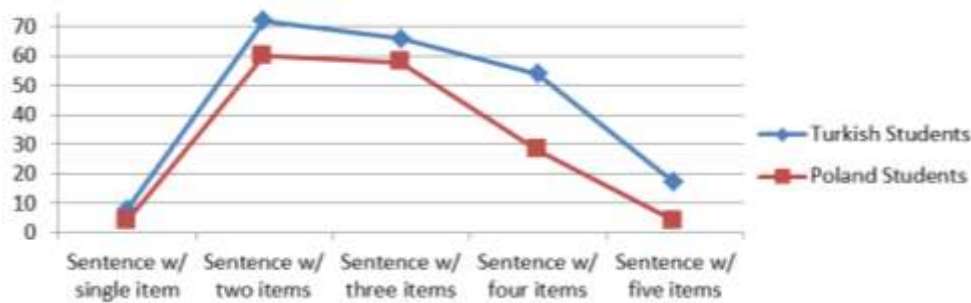
No meaningful difference in terms of word averages per sentences of students was found. It is seen that Polish students use average of  $\bar{X}=9.18$  while Turkish students use average of  $\bar{X}=10.20$  words Table 4. No meaningful difference in terms of sentence averages per essays of students was found. Therefore, Polish students tend to write essays with average of  $\bar{X}=15.7$ ; while Turkish students with average of  $\bar{X}=18.2$  sentences (Table 4).

### Suggestions

Data diversification and interpretation studies, which might offer source for distributional approach could be widened to other skill areas. Materials related to students'



**Graph 2.** The amount of sentences used by Turkish students in written communication.



**Graph 3.** Sentence setting types of Poland and Turkish students in written communication in terms of item similarities.

writing and speaking skills could be developed in order to eliminate the differences about setting similar sentences in syntax.

During writing lectures, it is suggested that more studies should be applied to students as giving them opportunities to create sentences, which are created with 4 or 5 items, while paying attention to Turkish syntax frequencies. Similar studies may be done to use in educational materials and to provide data.

### Conflict of Interests

The author has not declared any conflicts of interests.

### REFERENCES

- Büyüköztürk Ş, Çakmak EB, Akgün ÖE, Karadeniz Ş, Demirel F (2008). *Research Methods*. Ankara: Pegem.
- Büyüköztürk Ş (2013). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem.
- Chomsky N (1965). *Aspects of the theory of syntax*, Cambridge: The M.I.T.
- Chomsky N (1966). *Syntactic Structures*, Paris: Mouton & Co.
- Cohen L, Manion L, Morrison K (2000). *Research Methods in Education (5th Edition)*. London: Routledge Falmer.
- Demirci K (2010). What we understand from the Deep Structure and Surface Structure Concept?, *Turk. Stud.* 5(4):291-304. <http://dx.doi.org/10.7827/TurkishStudies.1762>
- Demirel E (2011). *Court of Divanü Lugati't Turk in Syntax*, master thesis, Pamukkale University, Denizli.
- Friedman C, Kra P, Rzhetsky A (2002). Two biomedical sublanguages: a description based on the theories of Zellig Harris. *J. Biomedical Informatics* 35:222-235.
- Gay L (1987). *Educational research: competencies for analysis and application*. Columbus, Ohio: Merrill Publishing Company
- İşkin GS (1999). *Linguistic Principles in Language Learning*, Cumhuriyet University Faculty Theology J. 3(1). Retrieved from "www.eskidergi.cumhuriyet.edu.tr.
- Landauer T, Dumais S (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* 104(2):211-240. <http://dx.doi.org/10.1037//0033-295X.104.2.211>
- Matthews PH (1997). *Concise Oxford Dictionary of Linguistics*, Oxford & New York: Oxford University.
- Miller G, Walter C (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes.* 6(1):1-28. <http://dx.doi.org/10.1080/01690969108406936>.
- Pantel P (2005). Inducing ontological cooccurrence vectors. *Proceedings of the 43rd Conference of the Association for Computational Linguistics*. Association for Computational Linguistics. 125-132.
- Piits L (2013). Distributional Hypothesis: Words For 'Human Being' And Their Estonian Collocates, *Trames.* 17(67/62) 2:141-158.
- Redington M, Chater N, Finch S (1998). Distributional Information: A Powerful Cue For Acquiring Syntactic Categories, *Cognitive Sci.* 22(4):425-469. [http://dx.doi.org/10.1016/S0364-0213\(99\)80046-9](http://dx.doi.org/10.1016/S0364-0213(99)80046-9)
- Sahlgren M (2008). The distributional hypothesis, *Rivista di Linguistica.* 20(1):33-53.
- Tura S (1983). Place of linguistics in language teaching, teaching the Turkish language, 9, Retrieved from <http://turkoloji.cu.edu.tr/DILBILIM/tura.pdf>.
- Turney D, Pantel P (2010). From frequency to meaning: vector space

- models of semantics, *J. Artificial Intelligence Res.* 37:141-188.
- Vardar B (1999). *XX. Century Linguistics*, Istanbul: Multilingual Publishing.
- Weeds J, Dowdall J, Schneider G, Keller B, Weir D (2005). Using distributional similarity to organise biomedical terminology, Retrieved from <http://users.sussex.ac.uk/~davidw/resources/papers/set-weeds-dowdall.pdf>
- Yildirim A, Şimşek H (2008). *Qualitative research methods in the social sciences*. Ankara: Outstanding.
- Zellig H (1968). *Mathematical structures of language*. Interscience.
- Zellig H (1970). *Distributional structure*. In *Papers in Structural and Transformational Linguistics*. University of Pennsylvania, USA.