



Dil Eğitimi Araştırmaları İçin Bir Değişken Önerisi: Kelime Hazinesi Katsayısı

*Özay KARADAĞ**

Öz

Toplum ve insani bilimlerle ilgili araştırmalarda bireyden bireye ya da gruptan gruba değişebilen özellikler ölçülerek sosyal olgular açıklanmaya çalışılır. Bireyler söz konusu olduğunda dil, en temel özelliklerden biri olarak karşımıza çıkmaktadır. Dil performansının ölçülmesi ve bunun üzerinden oluşturulabilecek indeksler, bireyin diğer özellikleriyle kolaylıkla ilişkilendirilebilir. Bu özelliklerden biri de sahip olunan kelime hazinesidir. Bireylerin kelime hazinesinin tam olarak belirlenmesi boyutsal ve çok zahmetli bir çalışmayı gerektirir. Bu yönüyle örneğin nicel bir araştırmada oluşan bir örneklemin her biri için uzun süreli kelime hazinesi araştırması gerçekleştirmek mümkün değildir. Bu durum kelime hazinesi düzeyini gösterecek başka bir çözümü zorunlu kılmaktadır. Bir kişinin herhangi bir anlatımında kullandığı farklı kelimelerin toplamda kullandığı kelime sayısına bölünerek elde edilen **Kelime Hazinesi Katsayısı** değeri bireysel kelime hazinesi zenginliğini betimlemede önemli bir ipucudur. Bu değer kelime hazinesinin kullanımına ilişkindir ve birçok başka özellikle ilişkilendirilebilir. Bu çalışmada dil eğitimi araştırmalarında **Kelime Hazinesi Katsayısı**'nın nicel bir değişken olarak nasıl kullanılacağı ele alınmıştır.

Anahtar Kelimeler: Değişken, kelime hazinesi katsayısı, dil eğitimi, nicel araştırma.

A Variable Suggestion for Language Education Research: Vocabulary Coefficient

Abstract

Social and humanities research are tried to explain the social phenomena by measuring characteristics that can be changed by individuals or groups. Language -when the individual is the subject matter- emerges as one of the most basic features. The measurement of language performance and the indexes that can be built through it can easily be related to other characteristics of the individual. One of these features is the vocabulary. The precise identification of the vocabulary of individuals requires longitudinal and very laborious work. In this respect, it is not possible to carry out a long-term vocabulary search for each sample of a quantitative study.. This situation requires another solution that will show the level of vocabulary. The Vocabulary Coefficient value obtained by dividing the number of words a person uses in any given expression into the total number of words used is an important clue in describing the richness of individual vocabulary. This value is related to the use of vocabulary and can be associated with many other features. This study deals with how to use the Vocabulary Coefficient in quantitative language education researches as a variable.

Keywords: Variable, vocabulary coefficient, language education, quantitative research.

* Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Ankara, ozaykaradag@hacettepe.edu.tr, ozaykaradag@gmail.com.

Giriş

Dil düzeyinin temel göstergelerinden biri, anlatımda kullanılan kelime sayıdır. Herhangi bir dilde sözlü ya da yazılı anlatım yapan bireyin kullandığı kelime sayıları iki özellik gösterir:

1. Farklı kelime sayısı: Anlatım içinde birbirinden farklı sözlük birim olarak kullanılan kelimelerin sayısını ifade eder.

2. Toplam kelime sayısı: Anlatım içinde tekrara bakılmaksızın kullanılan toplam kelime sayısını ifade eder.

Kelime hazinesi araştırmaları genellikle sıklık çalışmaları olarak gerçekleştirilmektedir (Baş, 2006; Karadağ, 2005; Kurudayıoğlu, 2005). Bu tip çalışmalarda kullanım sıklığı belirlenerek, en sık kullanılan kelimedenden en düşük sıklıkta kullanılan kelimeye doğru sıralanmış kelime listeleri oluşturulur. Bunun yanı sıra, farklı kelime sayısı ve toplam kelime sayısı elde edilerek bu değerler kelime hazinesine ilişkin genel değerlendirmelerde kullanılır.

Sıklığa dayalı kelime hazinesi araştırmalarının en önemli özelliği, büyük metin hacimlerinin analizi olarak gerçekleştirilmesidir. Bu tip çalışmalarda metin havuzlarının genişliği, araştırma sonuçlarının güvenilirliğini belirlemektedir. Bir kişi ya da grup için yapılan çalışmalarda elde edilebilecek en büyük metin hacmine ulaşılmaya çalışılması bu araştırmaları zaman, maliyet ve emek bakımından güçleştirmektedir. Ayrıca elde edilen sonuçlar standartlaştırılmadığından bu sonuçların nicel araştırmalarda bir değişken olarak kullanılması da söz konusu olmamaktadır.

Mevcut kelime hazinesi araştırmalarında tek tek bireylerin kelime hazinesi büyüklüğünü belirleyerek standart bir değer üretmek ve kelime hazinesi ile bireyin diğer özellikleri arasında ilişki açıklamak, böylelikle ulaşılan sonuçları genellemek mümkün olmamaktadır. Oysa bireyin kelime hazinesi özelliğinin eğitim araştırmalarında kullanılan birçok değişkenle ilişkili olabileceği unutulmamalıdır. Bu nedenle bireylerin kelime hazinesi ile cinsiyet, sosyo-ekonomik düzey, okuduğunu anlama başarısı, dinlediğini anlama başarısı, akademik başarı, strateji kullanma düzeyi vb. arasında ilişkileri açıklamak için yeni bir değişkene ihtiyaç duyulmaktadır.

Bu bağlamda bireylerin kelime hazinesi özelliği ile diğer özellikleri arasında ilişkileri açıklayabilmek için kelime hazinesi özelliğinin standart bir değer olarak belirlenmesi gerekir. Kelime hazinesine ilişkin oluşturulacak indeks başta dil eğitimine yönelik olmak üzere eğitim araştırmalarında bir değişken olarak kullanılabilir.

Bir Değişken Olarak Kelime Hazinesi Katsayısı

Toplumsal araştırmalarda sonuca ulaşmak için değişken analizi etkili bir yoldur. Bu bakımdan değişken kavramı eğitim araştırmaları açısından büyük önem taşır. Bir araştırmada *araştırmanın*

Dil Eğitimi Araştırmaları İçin Bir Değişken Önerisi: Kelime Hazinesi Katsayısı

olgularından oluşan, birden fazla değer alabilen; değişebilen her şey (Aziz, 1994) veya bir kişinin veya birimin değeri değişebilen özelliği (Scott ve Morrison, 2016) şeklinde tanımlanan değişken, bir durumdan diğerine farklılık gösteren nitelikler (Turgut ve Baykurt, 2012) olarak da ifade edilebilir.

Sıklığa dayalı kelime hazinesi araştırmalarında bir grubun ya da bireylerin kullandığı kelimeler belirlenirken sayısal olarak da tek tek kelimelerin kullanım sıklığı, sözlük birim olarak kullanılan farklı kelime sayısı ve toplam olarak kullanılan kelime sayısı değerleri elde edilmektedir. Bu sayısal değerler, bireylerin veya grupların sözlü (transkripte edilerek) ya da yazılı anlatımlarının bir metin havuzu hâline getirilip çözümlenmesiyle oluşturulmaktadır.

Bir grubun ya da bireylerin her birinin kelime hazinesine ilişkin sayısal değerlerden ikinci bir veri de kullanılan farklı kelime sayısının toplam kelime sayısına bölünmesiyle elde edilmektedir ve bu yolla elde edilen değer **Kelime Hazinesi Katsayısı** olarak adlandırılmaktadır (Baş, 2006; Gündüz, 1996; Karadağ, 2005; Kurudayıoğlu, 2005). Bu değer 0 (sıfır) ile 1 (bir) arasında değişmektedir. Bir metindeki kelimelerin her biri farklı kelime (farklı bir sözlük birim) ise değer 1 (bir) olmaktadır. Örneğin 100 kelimedenden oluşan bir metin 50 farklı kelime ile oluşturulmuşsa $50/100= 0,5$ değeri elde edilir. Bu değer metindeki her iki kelimedenden birinin farklı kelime olarak kullanıldığını göstermektedir.

Kelime Hazinesi Katsayısının değeri sayısal olarak ifade edildiği ve standartlaştırılabildiği için dil eğitime yönelik araştırmalarda nicel bir değişken olarak kullanılabilir. Stevens'in (1946) sınıflamasına göre bu indeks **sürekli** olup **aralıklı** bir değişkendir. Bu değişken 0 (sıfır) ile 1 (bir) arasında sonsuz sayıda değer alabileceğinden **sürekli**, 0 (sıfır) noktası ile değişkenin 0 (sıfır) değerinin birbirine karşılık olmaması dolayısıyla **aralıklıdır**.

Kelime Hazinesi Katsayısı yükseldikçe bireyin, metnin ya da metin havuzunun kelime hazinesinin zenginliğinden bahsedilebilir. Ancak toplam kelime kullanımı -yani hacim- arttıkça bu değer düşmektedir. Oysa toplam kelime kullanımına karşılık gelen hacmin artması da dil kullanımının, dolayısıyla kelime hazinesinin zenginliğine işaret eden önemli göstergelerdendir. Böylesi bir durumda **Kelime Hazinesi Katsayısı** değerinin düşmesine sebep olan unsurlardan biri, örneğin dilde iletişimsel olarak kullanılan "**bir, ve, ile, gibi, kadar vb.**" kelimelerin kullanım sıklığının artması, bir başka deyişle incelenen metinde ya da metin havuzundaki tekrarının yükselmesidir.

Genel olarak tekrarlar arttıkça kelime hazinesi katsayısı düşmektedir. Bu yüzden bir incelemede iki farklı grup, birim ya da tek tek bireylerin kelime hazinesi özelliklerini, kelime hazinesi katsayısı üzerinden karşılaştırabilmek için her birinin kelime hazinesi katsayısı değerinin eşit metin hacimlerin analizinden elde edilmiş olması gerekmektedir. Örneğin bir uygulamada 4 öğrenci sırasıyla öğrenciler 90, 100, 115, 120 kelimedenden oluşan metinler yazmış olsun. Bu dört öğrencinin karşılaştırılmasında **Kelime Hazinesi Katsayısı** değişkenin kullanılması için 90 kelime metin yazan

öğrencinin metin hacmi esas alınarak diğer öğrencilerin metinlerinden ilk 90 kelime değerlendirmeye alınmalıdır. Her öğrenci için ayrı ayrı olmak üzere toplam 90 kelime içinde kaç farklı kelime kullandığı belirlenmeli her öğrencinin kullandığı farklı kelime sayısı 90'a bölünerek her bir öğrenci için **Kelime Hazinesi Katsayısı** elde edilmelidir.

Araştırmalarda bireylerin her birinin **Kelime Hazinesi Katsayısını** el elde etmek için yazılı ya da sözlü anlatımlardan elde edilen metinlerin toplam kelime sayısının alt sınırının ne olacağı belirlenmelidir. Herhangi bir araştırmada kelime hazinesi katsayısı değeri üretmeye yönelik verilen bir anlatma yönergesini, örneğin her biri farklı üç kelimedenden oluşan bir cümle ile gerçekleştiren bir öğrencinin bu değeri $3/3=1$ olacaktır. Elde edilen bu sonuç olabilecek en yüksek **Kelime Hazinesi Katsayısı** değeri olmakla birlikte anlatımda kullanılan kelime sayısının 3 ile sınırlı olmasından dolayı gerçekliği yansıtmayacaktır. Karadağ'ın (2005: 61-62) çalışmasındaki 70 farklı ilkökul sınıfındaki metin uzunluklarının ortalamalarının toplamı 4551'dir. Bu değer sınıf sayısına bölüldüğünde $4551/70=65$ sonucu elde edilmektedir. Birinci ve ikinci sınıf düzeyinde bu ortalamanın daha düşük olduğu (Karadağ, 2005: 61-62) dikkate alınarak metin uzunluklarının alt sınırı 50 kelime olarak belirlenebilir. Bu durumda toplam kelime sayısı 50'den az olan metinlerde **Kelime Hazinesi Katsayısı** hesaplanmamalıdır.

Bir araştırmada, örneğin 300 kişilik bir öğrenci grubuna bir konu verilerek onların yazılı ya da sözlü anlatım yapmaları istenir. Öğrencilere ait metinler uzunlukları/toplam kelime sayıları bakımından değerlendirilerek 50 kelimedenden daha kısa olan örneğin 10 metin değerlendirme dışı bırakılır. Kalan 290 öğrenciye ait metinlerin toplam kelime sayılarının 55 ile 120 kelime arasında değiştiğini varsayalım. Bu durumda 55 kelimelik metin esas alınarak diğer öğrencilerin metinlerinin ilk 55 kelimesi değerlendirilir. Her öğrencinin 55 toplam kelime içinde kaç farklı kelime kullandığı ayrı ayrı belirlenir. Öğrencilerin her birinin kullandığı farklı kelime sayısı toplam kelime sayısı olan 55'e bölünerek yine her öğrenci için **Kelime Hazinesi Katsayısı** elde edilmiş olur. Bu değerler öğrencilerden elde edilen diğer değişken değerleri ile ilişkilendirilerek değişken analizinde kullanılabilir.

Sonuç

Kelime Hazinesi Katsayısı standartlaştırılabilir bir sayısal değerdir ve bu yönüyle değişken analizlerinin yapıldığı nicel araştırmalarda değişken olarak kullanılabilir. Bir anlatımda kullanılan farklı kelimelerin toplam kelimelere oranını ifade eden **Kelime Hazinesi Katsayısı** indeksi nicel nitelikli dil eğitimi araştırmalarında kullanılabilceği gibi sosyal bilimlerin diğer alanlarında da kullanılarak araştırmacıları genellenebilir sonuçlara ulaştırabilir. Örneğin, yapısal eşitlik modelleri oluşturmada kullanılabilir.

Dil Eğitimi Araştırmaları İçin Bir Değişken Önerisi: Kelime Hazinesi Katsayısı

Kaynaklar

- Aziz, A. (1994). *Araştırma yöntemleri-teknikleri ve iletişim*. Ankara: Turhan Kitabevi.
- Baş, B. (2006). *1985-2005 yılları arasında çocuk edebiyatı sahasında yazılmış tahkiyeli metinlerin söz varlığı üzerine bir araştırma*. (Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü Yayınlanmamış Doktora Tezi). Ankara.
- Gündüz, M. (1996). *Basın ve terör*. İzmir: Saray Kitabevleri.
- Karadağ, Ö. (2005). İlköğretim I. kademe öğrencilerinin kelime hazinesi üzerine bir araştırma. (Gazi Üniversitesi Eğitim Bilimleri Enstitüsü. Yayınlanmamış Doktora Tezi). Ankara.
- Kurudayıoğlu, M. (2005). İlköğretim II. kademe öğrencilerinin kelime hazinesi üzerine bir araştırma. (Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü Yayınlanmamış Doktora Tezi). Ankara.
- Scott, D. ve Morrison, M. (2016). *Eğitim araştırmasında temel fikirler ve kavramlar*. (Çev.: Ümit Tatlıcan) İstanbul: Sentez Yayıncılık.
- Stevens, S.S.(1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Turgut, M.F. ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi Yayınları.

Extended Abstract

Introduction

One of the key indicators of language level is the number of words used in narration. The number of words used by a person in spoken or written language on any level has two characteristics:

1. Number of different words: Expresses the number of words used in different lexical units.
2. Total number of words: Expresses the total number of words used irrespective of repeating.

Vocabulary researches are mostly carried out as frequency studies. In this type of study, at first the frequency of usage is determined and then word lists from the most commonly used words to the least frequently used words are created. Besides, these values are used in general evaluations about vocabulary by obtaining different word counts and total word counts.

One of the most important features of word-based research is the analysis of large text volumes. The width of text pools in such studies determines the reliability of research results. Trying to reach the greatest volume of text in such studies for a person or group makes these researches difficult regarding to time, costs and labour spent on them. Moreover, since the results obtained are not standardized, it is not possible to use these results as a variable in quantitative educational research.

In available vocabulary researches, it is not possible to produce a standard value by determining the size of vocabulary of individuals one by one and also to establish a relationship between vocabulary and other characteristics of the individual so that the results reached can not be generalized. It can be argued, however, that the vocabulary of an individual may be related to many variables used in educational research. It is possible to found a relation among individual's gender, socio-economic level, successful reading comprehension, understanding success, academic achievement, level of strategy use, etc.

In order to be able to establish relationships between the vocabulary and other characteristics of individuals, the vocabulary feature should be defined as a standard value. The index to be formed for vocabulary can be used as a variable in educational research, especially for language education.

Vocabulary Coefficient as a Variable

Variable analysis is an effective way to reach the conclusion in social research. In this respect, the concept of variable is of great importance in terms of educational research. In a research, the variable that can be described as; qualities show difference from one situation to another (Turgut and Baykurt, 2012), a variable that is defined as something that can change (Aziz, 1994) or a person or a unit whose value can be changed (Scott and Morrison, 2016).

While the words used by a group or individuals are determined in the word-based searches, the number of words used as a dictionary unit, the number of words used as a dictionary unit, and the number of words used as a total are obtained numerically. These numerical values are formed by interpreting a textual pool of verbal (transcribed) or written expressions of individuals or groups.

A second data from the numerical values of a group or individual's vocabulary is obtained by dividing the number of different words used by the total number of words and the value obtained in this way is called the Vocabulary Coefficient (Gündüz, 1996, Karadayıoğlu, 2005; Chief, 2006). This value ranges from 0 (zero) to 1 (one). If each word in a text is a different word (a different dictionary unit), the value is 1 (one). For example,

if a text consisting of 100 words is composed of 50 different words, $50/100 = 0.5$ is obtained. This value indicates that one word out of every two words in the text is used as a different word.

Since the value of the Vocabulary Coefficient can be numerically expressed and standardized, it can be used as a quantitative variable in research on language teaching. According to Stevens (1946) classification, this index is continuous and intermittent. Because this variable can have infinite number of values between 0 and 1, it is continuous because the 0 (zero) point and the 0 (zero) value of the variable do not correspond to each other.

As the Vocabulary Coefficient increases, the wealth of the individual, the text, or the vocabulary of the text pool can be mentioned. However, as the total word usage - the volume - increases, this value decreases. However, the increase in vocabulary corresponding to the total vocabulary usage is also an important indicator of the use of language, hence the richness of vocabulary. In such a case, one of the factors that cause the value of the Vocabulary Coefficient to decrease is the increase in the frequency of use of words such as, for example, and so on, which is used communicatively on the dictation, in other words the repetitions in the text or text pool? In general, as the number of repetitions increases, the coefficient of vocabulary decreases. Therefore, in order to be able to compare the vocabulary characteristics of two different groups, units or individuals in a review, the vocabulary coefficient value of each must be obtained from the analysis of equal text volumes. For example, in an application, 4 students have written texts consisting of 90, 100, 115, 120 words respectively. In the comparison of these four learners, the first 90 words from the texts of the other students should be evaluated based on the text volume of the student who writes 90 words in order to use the Vocabulary Coefficient variable. For each student, the number of different words used in a total of 90 words should be divided by 90 for each student and the Vocabulary Coefficient should be obtained for each student.

In research, to determine the Vocabulary Coefficient of each of the individuals, it is necessary to determine what the lower limit of the total number of words of the texts obtained from written or verbal expressions is. In a given study, a student who performs a given instruction to produce a value of a vocabulary coefficient value, for example, each with a sentence consisting of three different words, will have this value $3/3 = 1$. This result can not be a reality because the number of words used in the narration is limited to 3, with the highest possible Coefficient of Vocabulary. In Karadağ's (2005: 61-62) study, the total numbers of the average lengths of texts in 70 different elementary school classes is 4551. When this value is divided by the number of classes, $4551/70 = 65$ results are obtained. Considering that this average is lower at the first and second grade level (Karadağ, 2005: 61-62), the lower limit of text lengths can be specified as 50 words. In this case, the Vocabulary Coefficient should not be calculated for texts with a total of less than 50 words.

In a research, for example, a group of 300 students are asked to give a written or oral explanation by giving them a subject. The texts of the students are evaluated in terms of length / total word counts, and 10 texts shorter than 50 words are excluded from the evaluation. Considering the total number of words in the remaining 290 student texts varies between 55 and 120 words. In this case, the first 55 words of the texts of the other students are evaluated based on the 55-word text. How many different words each student used in 55 total words is determined separately. The number of different words used by each student is divided by the total number of words, 55, and the Vocabulary Coefficient for each student is obtained. These values can be used in the analysis of variables by associating them with other variable values obtained from the students.

Result

The Vocabulary Coefficient is a standardized numerical value, which can be used as a variable in quantitative social studies in which variable analyses are made. Researchers can make researchers reach generalizable results by using The Vocabulary Coefficient Index, which expresses the ratio of different words to total words used in a narrative can make the researchers reach generalizable results by using it in a quantitative language education research or in other fields of social sciences. For example, it can be used in forming Structural Equation Models.