# A Corpus Based Study on Gerundials: Finding the Primary Gerundial Structures for Teaching Turkish as a Foreign Language

**Bülent Özkan**
*Mersin University (Turkey)*
ozkanbulent@gmail.com

## Abstract

*Nowadays, linguistic studies can be conducted in databases in which research questions have the possible answers, and those answers of the research questions can provide descriptions. Gerundial, which provides expressive functions, prevents redundant use of the words and makes the thought easy to transfer, is important in Turkish regarding the frequency of use. Up to now, when the studies on gerundial have been taken into consideration, it can be claimed that previous studies have covered a limited area of Turkish archives. In that sense, it is of high importance to evaluate Turkish with a corpus to teach Turkish as a foreign language. Therefore, the aim of the present study is to evaluate gerunds, participles, verbal adverbs, which are in the category of verbals/gerundial for many researchers, with a corpus of 35.000.000 words (+/-) to find out the primary structures for teaching Turkish as a foreign language. In this respect, the research question of the study is "What are the primary gerundials in teaching Turkish as a foreign language?" This study has been prepared in line with the data of the project (number: 114R033) of The Scientific and the Research Council of Turkey named as "Gerundials in Turkey Turkish"*

## 1. Introduction

Corpus can be defined as the compile of texts obtained from oral and written data of a particular language in accordance with the standards, and the aim of corpus is to provide examples with the help of concrete linguistic criteria.[1] Corpora are accepted as fast and reliable platforms which provide data, and they have their own spesific techniques which are widespread in linguistics. Particularly, corpora, which are used as research samples in order to clarify the prior structures for language teaching, are fruitful for providing scientific data about language.
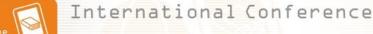
Nowadays, corpora prepared in accordance with corpus linguistic principles and techniques contribute a lot to the literature in that they provide experimental results related to the research questions. Therefore, the present study aims to reveal the priorities, usage frequencies and functions of gerundials in Turkey Turkish. The reason for conducting research on gerundials via a corpus is that gerundials make Turkish eloqeunce, prevent unnecessary repetitions, provide easy access for transferring thoughts, and connect sentences fluently and easily. [2], [3], [4], [5], [6] In that sense, the aim of the study is to identify these issues via the verb "gel-" which is the second common item in the Turkish corpus regarding its usage frequency. In line with this purpose, the research question of the study is "What are the primary gerundials in teaching Turkish as a foreign language?"

Besides the issues mentioned above, the study has also some limitations. To ilustrate, rather than using all gerundials used in previous studies, [7], [8] the verb "gel-", which is the second common verb among 6441 verbs, will be examined in line with its use and its usage frequencies.

## 2. Methodology

This research has been designed in line with descriptive and correlational models. Descriptive research attempts to describe the things as they are, while correlational research is the research type that does not present reason-result relationship. This type of research cannot be controlled or can be controlled because of its nature.[9], [10], [11] Regarding data collection tools, some researches apply methods, tools and techniques prepared in line with spesific purposes.[9] Within this scope, the present research applies the methods and techniques corpus linguistics provides.

In order to collect data, Turkish Corpus II (including 35 million words +/-), created via the project of 109K516 - The Scientific and Technological Research Council of Turkey - SOBAG The Control of Verbs' Corpus and Corpus Based Dictionary in Turkish Vocabulary, is going to be used as data collection tool. [7], [8]

## 2.1. The Content of Turkish Corpus II

Turkish Corpus II has three layers regarding its ability to represent, sampling, and distributional properties in this research. In the first layer, literary books are included. Moreover, it has been considered that this layer has the oppurtunity to represent Turkish Literary Language in a very good way. The second layer included newspaper texts and journals published in 2009, and the third layer includes (as indicated in table below) books which have been digitized because of word limit.

When it is thought under this perspective, Turkish Corpus II is a literary language corpus, and it does not include oral language. When the texts are considered in line with timeline, it can be said that the first layer includes texts about 80 % from the years 1970-2009; about 12 % from the years between 1950-1970; and, 8 % before 1950s. On the other hand, the second layer includes texts from the year 2009 totally, and the texts of the third layer have been chosen between the years 1990-2009 (see Table 1).

Turkish Corpus II mentioned above has 35.000.000 words (+/-), 810.000 singular words (+/-).Turkish Corpus II has been parsed. For every sentence, main layers and sub layers of every sentence have been tagged as out of text parts. After this process, lemmatizing has been done and words have been deduplicated. [7], [8]

*Table 1. Layers, text types, number of sentences and time interval in Turkish Corpus II*

| LAYERS | NUMBER OF SENTENCES | | % | TIME INTERVAL |
|---|---|---|---|---|
| **1. LITERARY TYPES** | | | | |
| Novel | 568.722 | 38,37 | | |
| Story | 220.967 | 23,98 | | |
| Criticism | 144.689 | 38,12 | | |
| Theater | 111.811 | 4,02 | | |
| Examination-Research | 91.386 | 6,17 | | REPUBLICAN PERIOD TURKISH LITERATURE (1920-2009) |
| Memory | 81.624 | 5,51 | | |
| Conversation-Article | 69.412 | 4,68 | | |
| Trip | 52.891 | 3,57 | | |
| Poem | 47.092 | 3,18 | | |
| Humor | 40.606 | 2,74 | | |
| Biography | 36.933 | 2,49 | | |
| Layer | 16.028 | 1,08 | | |
| **Total number of the layers** | **1.482.161** | | **53,25** | |
| **2. INTERNET TEXTS** | | | | |
| News, economy, politics, strategy, life, popular science etc. (newspaper and journals) | | | | 2009 |
| **Total number of the layers** | **921.488** | | **33,11** | |
| **3. VARIOUS TYPES** | | | | |
| eating habits, computer science, biology, behavioral sciences, linguistics-semiotics, legend, educational sciences, ecology, information, aesthetics, evolotion, philosophy, narratice, physics, geometry, law, financial sciences, statistics, chemistry, personal development, logic, story, maths, architecture, mythology, engineering (various), music, speech, psychology, painting, art, cinema, sociology, strategy, history (various), medicine-health, zoology, geology, child literature (various) | | | | 1990-2009 |
| **Total number of the layers** | **379.528** | | **13,64** | |
| **TOTAL** | **2.783.177** | | **100,00** | |

## 2.2.Syntactical Tags for Gerundials

In order to create some spesific features of the corpus, the syntactical corpus tagging tool developed by Kemal Oflazer[12] has been used, and lemmatized have been tagged in line with their specific features.[13], [14] In this context, morphological taggers for the gerundials in Turkish are illustrated in Table 2 below.

*Table 2: Special features of the taggers that belong to gerundials*

| Root | Derivations | | Taggers of Gerundials |
|---|---|---|---|
| **Verb** | **Polarity** | **+Pos**<br>**+Neg** | **+ParsPart** (-m$^O$ş, -D$^O$k)<br>**+FutPart** (-AcAk) |
| | **Compounding-Modality** | **+Able** (-ebil)<br>**+Repeat** (-edur)<br>**+Hastily** (-ıver)<br>**+EverSince** (gel-)<br>**+Almost** (yaz-)<br>**+Stay** (kal-)<br>**+Start** (koyul-) | **+PresPart** (-An, -mAz)<br>**+AfterDoingSo** (-$^O$p)<br>**+SinceDoingSo** (-AlI)<br>**+As** (-D$^O$kÇA)<br>**+When** (-$^O$ncA)<br>**+ByDoingSo** (-ArAk)<br>**+While** (-ken)<br>**+AsIf** (-cAsInA, -AsIyA)<br>**+WithoutHavingDoneSo** (-mAdAn)<br>**+Inf** (-mA, -mAk, -$^O$ş) |

By means of the taggers used for gerundials (see Table 2), the frequencies of the data have been identified. For this identification, the negative positive use of the gerundials and compound verbs have also been listed.

## 3. Findings and Discussion

As a result of corpus queries, it has been found out that subcategories regarding the use of gerundials are listed as participles, gerundials and adverbials. In that sense, total frequencies are displayed in Table 3.

*Table 3: Sub categories of the gerundials*

| Type of the Gerundial | Participles | Gerundials | Adverbials |
|---|---|---|---|
| **General Frequency** | 73.626 | 20.003 | 16.112 |

When considered from this point of view related to Turkish texts, it can be said that participles need to be attached importance. Sub categories of the gerundials are discussed in sections 3. 1, 3. 2, and 3. 3.

## 3.1. Participles

When the most frequent use of participles are examined in Table 4 below, it can be seen that **+PresPart** *(-An)*, **+ParsPart** *(-d$^O$k)*, **+FutPart** *(-AcAk)*, **+ParsPart** *(-m$^O$ş)*, **+PresPart** *(-mAz)* are the most common ones. In addition to this, it has been found out that the usage density of **(+Pos)** decline is more common than **(+Neg)** decline and all other declines.

**+PresPart** *(-mAz)* is frequent because it is included in *(-ir / -irmez, gelir gelmez)*. Thus, it can be said that participles illustrated with an example verb "gel-" in Table 4 are of high importance while teaching Turkish as a foreign language.

*Table 4: The frequency distribution of participles*

| Verb | Taggers | +Pos | +Neg | Able+Pos | Able+Neg | +Hastily |
|---|---|---|---|---|---|---|
| **gel-** | **+PresPart (-An)** | 29.859 | 537 | 50 | 45 | 36 |
| | **+ParsPart (-d$^O$k)** | 25.125 | 1.427 | - | 174 | 6 |
| | **+FutPart (-AcAk)** | 11.993 | 1.210 | 1.240 | 297 | 24 |
| | **+ParsPart (-m$^O$ş)** | 1.560 | 16 | - | - | - |
| | **+PresPart (-mAz)** | 19 | 8 | - | - | - |

## 3.2. Gerundials

When the most frequent use of gerundials are examined in Table 5 below, it can be seen that **+Inf** *(-mA)*, **+Inf** *(-$^O$ş)*, **+PresPart** *(-mAktA)*, **+Inf** *(-mAk)* are the most common ones. Furthermore, when their use is examined in Table 5 below, it can be said that the density of the usage density of **(+Pos)**

decline is more common than **(+Neg)** decline and all other declines **(+Neg)**. As a result of corpus queries, it can be said that following gerundials in Table 5 may be of prior importance.

*Table 5: The density frequencies of gerundials*

| Verb | Taggers | +Pos | +Neg | Able+Pos | Able+Neg | +Hastily |
|------|---------|------|------|----------|----------|----------|
| | +Inf (-mA) | 8.113 | 363 | 96 | 11 | - |
| gel- | +Inf (-$^O$ş) | 7.017 | 29 | - | 2 | - |
| | +PresPart (-mAktA) | 2.995 | 154 | 10 | 10 | - |
| | +Inf (-mAk) | 1.061 | 86 | 55 | | 1 |

### 3.3. Adverbials

Taking into consideration the use of adverbials in Turkish, the most common adverbials have been found out as **+AfterDoingSo** (-$^O$p), **+When** (-$^O$ncA), **+ByDoingSo** (-ArAk), **+While** (-mAdAn), **+While** (-ken), **+As** (-D$^O$kÇA), **+FutPart** (-A), **+SinceDoingSo** (-AlI), **+AsIf** (-CasInA, -sIyA) (see Table 6).

Table 6 also displays that the usage frequencies of **+AfterDoingSo** (-$^O$p), **+When** (-$^O$ncA) are extremely close to each other. Also, it can be seen that **(+Pos)** decline is more common than **(+Neg)** and all other declines, and **+FutPart** (-A) has a frequency because of duplicated structures. Therefore, it can be said that the priorities of the following adverbials in Table 6 need to be taken into consideration while preparing a teaching material.

*Table 6: The frequency distributions of adverbials*

| Verb | Taggers | +Pos | +Neg | Able+Pos | Able+Neg | +Hastily |
|------|---------|------|------|----------|----------|----------|
| | +AfterDoingSo (-$^O$p) | 5.943 | 14 | - | 1 | - |
| | +When (-$^O$ncA) | 5.076 | 85 | 3 | 5 | 6 |
| | +ByDoingSo (-ArAk) | 1.364 | 2 | - | 2 | - |
| | +While(-mAdAn) | 1.310 | 7 | - | - | - |
| gel- | +While (-ken) | 1.243 | 17 | 2 | 1 | - |
| | + As (-D$^O$kÇA) | 353 | 34 | - | 2 | - |
| | +FutPart (-A) | 412 | - | - | - | - |
| | +SinceDoingSo (-AlI) | 181 | 3 | - | - | - |
| | +AsIf (-CasInA, -sIyA) | 42 | 4 | - | - | - |

## 4. Results and Conclusion

In the light of the results mentioned above, it can be concluded that is important to identify prior structures for creating an effective and qualified teaching background. These identifications contribute a lot in that they are based on experimental applications rather than conceptual examplifications.[15] Under this discussion, the findings of the present study conducted on gerundials via the corpus findings will contribute a lot to teaching Turkish as a foreign language.

## Acknowledgements

## References

[1] McEnery, T. vd. 2006. *Corpus-Based Language Studies -An Advanced Resource Book-*. London: Routledge.
[2] Atabay, N. vd. 1981. *Türkiye Türkçesinin Sözdizimi*. Ankara: TDK Yayınları.
[3] Banguoğlu, T. 2000. *Türkçenin Grameri* (6. Basım). Ankara: TDK Yayınları.
[4] Gencan, T. N. 2001. *Dilbilgisi*. Ankara: Ayraç Yayınevi.
[5] Korkmaz, Z. 2003. *Türkiye Türkçesi Grameri (Şekil Bilgisi)*. Ankara: TDK Yayınları.
[6] Karahan, L. 1999. *Türkçede Söz Dizimi* (6. Basım). Ankara: Akçağ Yayınları.

[7]  Özkan, B. 2012. *Türkiye Türkçesi Sözvarlığında Fiillerin Derlem Denetimi ve Derlem Tabanlı Sözlüğü*. Tübitak Proje no: 109K516. Mersin

[8]  Özkan, B. 2013. "Yöntem ve Uygulama Açısından 'Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü'", *bilig/Türk Dünyası Sosyal Bilimler Dergisi*, 66, 149-178.(SSCI).

[9]  Erkuş, A. 2009. *Davranış Bilimleri İçin Bilimsel Araştırma Süreci* (İkinci Basım). Ankara: Şeçkin Yayıncılık.

[10] Karasar, N. 2009. *Bilimsel Araştırma Yöntemi*. Ankara: Nobel Yayın- Dağıtım.

[11] Büyüköztürk Ş. vd., 2010. *Bilimsel Araştırma Yöntemleri*. Ankara: PegemA Yayıncılık.

[12] (http://www.hlst.sabanciuniv.edu/TL/)

[13] Say, B. 2003. Türkçe için Biçimbirimsel ve Sözdizimsel Olarak İşaretlenmiş Ağaç Yapılı Bir Derlem Oluşturma. Proje No: 199E026. Proje Raporu-Ek-b: Türkçe Ağaç Yapılı Derlem Kılavuzu. Ankara.

[14] Oflazer, K. 1994. "Two-level description of Turkish morphology", *Literary and Linguistic Computing*, 9, 2.

[15] Özkan, B., Gündoğdu, A. E. 2011. "Ana Dili ve Yabancı Dil Olarak Türkçenin Öğretiminde Belirteçler -Derlem Tabanlı Bir Uygulama-". *Theoretical and Applied Researches on Turkish Language Teaching*. Editör: L. Uzun, Ü. Bozkurt. Essen/Almanya: Die Blaue Eule*.