

A Corpus-Driven Analysis of *-r* Dropping in Spoken Turkish

Yuji KAWAGUCHI

1. Introduction

As is well known, corpus linguistics has expanded to cover an increasing range of languages since the 1990s. Today, even in linguistic fields investigating the spoken variety of language, such as sociolinguistics and pragmatics, corpus data have an important role to play. In Turkish linguistics, the phonetic phenomenon of *-r* dropping has been a subject of empirical description for years, but to our knowledge, there has been no objective analysis based on spoken Turkish corpus. Examining corpus data often provides several new and relevant insights for linguistic phenomena already investigated. In this paper, we will first describe the corpus design of the Multilingual Spoken Corpora (MSC) Turkish. We will then follow the main course of our analysis of *-r* dropping in spoken Turkish. The relevancy of some factors that stimulate the dropping of the final *-r* will be examined in the present progressive suffix *-iyor*, and we will also analyze the correlation between the present progressive suffix and the word *bir* “one, certain.”

2. Spoken Turkish Corpora and *-r* dropping

2.1. MSC Turkish and Usage-Based Linguistic Informatics

The construction of the present corpora was originally conceived in the 21st Century COE Program “Usage-Based Linguistic Informatics”; see <http://www.coelang.tufs.ac.jp/english/index.html>.

We sent researchers and doctoral students to record spontaneous conversations; the following universities collaborated in this study: University of Aix-Marseille and University of Paris XIII for French, Autonomous University of Madrid for Spanish, Universiti Kebangsaan Malaysia for Malay, Tamkang University of Taiwan for Chinese, Moscow State University of Humanities for Russian. The series of MSC has been developed since 2004: French, Spanish, Russian, Malay, Turkish, Chinese and Japanese. For instance, the MSC French contains around 200,000 word tokens and the TUFSL Part Of Speech Search Engine (TUFSL-POSSE) of spoken French was published in 2008; see http://cblle.tufs.ac.jp/tag/fr/index_en.html.

As part of the 21st Century COE Program, I organized a research team with my Turkish colleagues of Marmara University and Istanbul University for the construction of MSC Turkish; for the outline of research and some examples, consult http://www.coelang.tufs.ac.jp/multilingual_corpus/tr/index.php?contents_xml=top&lang=en.

At the campus of Marmara University and at the Language Center of Istanbul, we recorded 27 free conversations. First, it is important to discuss the representativeness of the MSC Turkish. The spoken examples of Istanbul Turkish studied within our recordings represent a mixture of diverse varieties of contemporary Turkish in Turkey.

In fact, since people from different regions of the country come to live in Istanbul, it is a melting pot of various regional and sociolinguistic varieties of Turkish. In this sense, a kind of standard has been elaborated in this metropolis for mutual understanding among Turks from different regions. However, it should be noted that the blanket claim that Istanbul Turkish can be considered as the spoken standard of contemporary Turkish does not prevent the presence of inner variations in Istanbul Turkish.

In the framework of the 21st Century COE Program, most of the 27 spontaneous conversations were recorded in the campus of Marmara University, Istanbul in June 2005 and June 2006.¹ Considering the ‘simplicity’ of transcription, the 27 conversations are assumed to be turn-takings of two persons, i.e., dialogues.² With some control over the conversation material, the investigator can determine the topic the informants will talk about. However, the spontaneity of the produced speech will increase when informants are allowed to choose their own words.³ In most of the conversations, our informants chose topics on their own initiative. In the introductory part of some conversations, the informants attempted to obtain information about each other and thereafter proceeded to the main topic of conversation. In a few cases, the researcher prepared his predetermined subject to ask informants. Generally, our corpora do not involve a spontaneous speech about a predetermined topic, but are similar to spontaneous dialogues. The total length of the recordings is 10 hours 26 minutes and the total number of word tokens is around 93,000. Undoubtedly, this corpus is very small in comparison to the spoken corpora of English, German, or French. Nonetheless, in the context of spoken Turkish corpora, our corpus is probably the largest one, see Table 1.

Table 1. Recording Profile

	Corpus	Length (min)	Topic	Date
1	D101-03-05	28.77	Education and university	2005.6.23
2	D102	32.36	Cigarette	2005.6.23
3	D104	29.26	EU and Turkey	2005.6.23
4	D106	29.52	Television	2005.6.23
5	D107	35.42	On Turkey	2005.6.24
6	D108	22.46	Leisure	2005.6.24
7	D109	26.25	Hometown	2005.6.24
8	D110	29.12	Tourist spot	2005.6.24
9	D111	26.5	Vacation	2005.6.24
10	D112	28.18	Language education and Turkish	2005.6.24

¹ After the end of the 21st Century COE, the MSC Turkish project was succeeded by the Global COE Program ‘Corpus-based Linguistics and Language Education,’ <http://cblle.tufs.ac.jp/index.php?id=21>.

² We did not focus on the stylistic differences of speech data in Turkish in order to ensure that the construction of the corpora, such as text fragments that were read aloud, was not planned in advance.

³ Gibbon, Moore, and Winski (eds). 1998, p.103.

11	D201	5.40	Education as an occupation	2006.6.8
12	D202	38.32	Turkish culture	2006.6.21
13	D203	26.53	Test	2006.6.14
14	D204	26.53	Lecture	2006.6.14
15	D205	25.14	USA and Turkey	2006.6.21
16	D206	36.13	Students	2006.6.21
17	D207	18.38	Turkish literature and culture	2006.6.21
18	D208	30.54	Foreign language and Turkish	2006.6.21
19	D209	31.32	Poems	2006.6.29
20	D210	5.19	Leisure	2006.6.29
21	D211	15.28	Drams	2006.6.29
22	D212	26.53	Music	2006.6.29
23	D213	22.20	Vacation	2006.6.20
24	D214	5.51	Traveling abroad	2006.6.8
25	D215	24.13	French language	2006.6.8
26	D216	23.55	Istanbul	2006.6.8
27	D217	5.29	Spoken language	2006.6.8

Our 17 informants are 8 men and 9 women with ages ranging from 20 to 55 years. They are from different cities and towns of Turkey, including Ardahan, Elazığ, İstanbul, İzmir, and Kayseri. Most of them were undergraduate or graduate students or teachers. In other words, our informants can be considered as highly educated Turks.

Table 2. Informant Profile

		Sex	Age	Birth place	Profession
1	AHT	M	30	İstanbul	Teacher
2	AK	F	32	Elazığ	Graduate student
3	BG	F	26	İzmir	Researcher
4	BH	F	24	İstanbul	Graduate student
5	DH	F	26	Gebze	Teacher
6	EY	F	23	İstanbul	Student
7	FK	F	55	İstanbul	Teacher
8	HS	M	29	Kırşehir	Teacher
9	MG	M	28	Kayseri	Teacher
10	MÜ	F	40	İsparta	Student
11	NS	F	20	İstanbul	Student
12	ÖK	M	29	Bursa	Student
13	RŞŞ	F	38	Nevşehir	Teacher
14	SG	M	30	İstanbul	Teacher

15	ŞK	M	27	Sivas	Teacher
16	SY	M	36	Yozgat	Teacher
17	ÜD	M	30	Ardahan	Teacher

Based on the notations for describing dialogues defined by Morel and Danon-Boileau (1998: 5), which are designed specifically for prosodic and syntactic analysis, the following conventions in transcription have been adopted by Selim Yılmaz and Yuji Kawaguchi; see Table 3.

Table 3. Conventions in transcription⁴

(?), (!)	question, exclamation
-	start of discourse
#, ##, ###	short or long pause
ooo, aaa	emotional long vowel: e.g., yook “oh, no,” haayır “oh, yes”
(.)	hesitation or suspension
(.....)	dropped segment(s)
_____	accentuated part
e (eee)	filler
m (mmm)	confirmation
{.....}	Dislocation: e.g., hava güzel {bugün} = bugün hava güzel. “it is fine today”
[.....]	paralinguistic elements like smile, laugh, and cough
<.....>	overlapping
.....	relatively rapid part
xxx	unintelligible part
%.....%	special intonation or speech melody

The following fragment of dialogue is extracted from D107. The topic of conversation is Turkey. Dislocations that occur frequently in Spoken Turkish are enclosed in braces. For instance, the adverbial expression *en son* “most recently” in the first discourse of BH is dislocated and should be before *nereye* in a neutral word order: *En son nereye gittin?* “When did you go there most recently?” Overlaps are indicated by angle brackets. They occur quite frequently in the following verbal exchange between BH and MG: - <Bodru(.)> - <ben de Bodrum(.)> - Bodrum’a <gittim> - <Bod>rum’un neresindeydin(?) “- Bodru... - Me too, Bodrum. - I went to Bodrum. - In what part of Bodrum did you stay?” Paralinguistic elements are in square brackets: [gülme] “laugh.”

Example of D107

BH40- gidiyorum tabi %gitmez olur muyum(?) %ah% keşke gitsem {şimdi gene} [gülme]
 <nereye gittin {en son}(?)> hazır havalar sıcakken ### en son geçen sene Bodrum’a
 gitmiştim %güney saahillerine iniyoruz şimdi% [gülme]

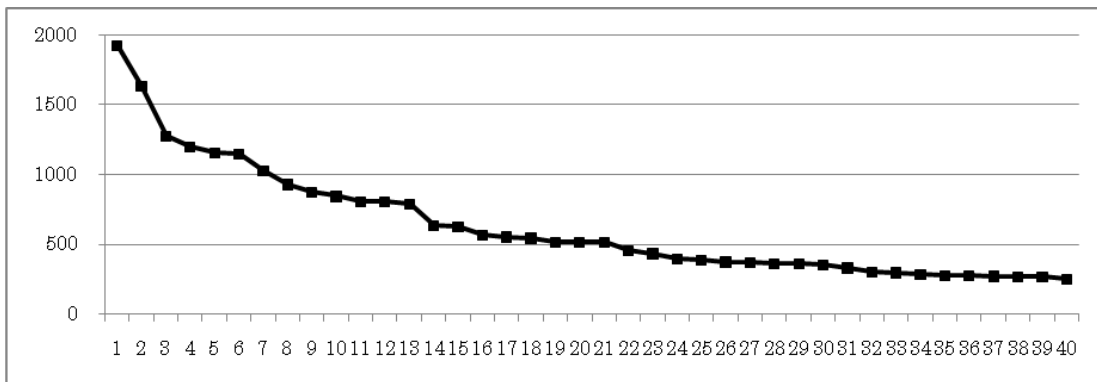
⁴ As for the details of corpus presentation, see Yılmaz (2006), pp. 202–205.

MG40- evet ####
 BH41- <Bodru(.)>
 MG41- <ben de Bodrum(.)>
 BH42- Bodrum'a <gittim>
 MG42- <Bod>rum'un neresindeydi(?)
 BH43- ## Gumbet

2.2. Frequency and word tokens

The frequency curve of word tokens is represented in the following graph. Occurrences of word tokens are indicated on the perpendicular line and frequency ranks on the horizontal line. From the 1st to the 9th token, the curve is very steep and followed by relatively equal occurrences from the 9th to the 13th token. There is a sharp drop between the 13th and 14th tokens and thereafter we can observe a progressive diminution of occurrences.

Figure 1. Frequency curve of word tokens



The top 20 words are given in Table 4. It is surprising to see that the most frequent word is the word *bir* without the final *-r*, while *bir* with final *-r* is the 6th on the list. This simple frequency analysis of word tokens demonstrates clearly that *-r* dropping is an extremely frequent phenomenon in spoken Turkish. However, its linguistic explanation needs to be refined on the basis of corpus evidence. For this reason, I insist on naming the present article a *corpus-driven analysis*.⁵

Table 4. Top twenty words

rank	word	occurrence
1	<i>bi(r)</i>	1928
2	<i>çok</i>	1639
3	<i>da</i>	1279
4	<i>yani</i>	1199

rank	word	occurrence
11	<i>ben</i>	805
12	<i>evet</i>	804
13	<i>var</i>	789
14	<i>ve</i>	631

⁵ For the term *corpus-driven*, see Teubert (2004), pp. 57–58.

5	<i>de</i>	1159
6	<i>bir</i>	1147
7	<i>bu</i>	1028
8	<i>o</i>	929
9	<i>ama</i>	875
10	<i>ee</i>	845

15	<i>e</i>	623
16	<i>şey</i>	567
17	<i>eee</i>	552
18	<i>için</i>	543
19	<i>işte</i>	514
20	<i>daha</i>	512

Most of the manuals of Turkish phonetics and phonology explain the *-r* dropping phenomenon. For example, in her article *Türkçe'nin Yapısı - I Sesbilim* A. Sumru Özsoy mentioned *-r* dropping in the section on consonant dropping.

8.2.2. Ünsüz Düşmeleri (Consonant dropping)

...

(i) /r/ düşmesi (/r/ dropping)

Bir sözcüğünün ve {-Iyor} biçimbiriminin son sesi konumunda bulunan /r/ sesi hızlı konuşmada çoğunlukla düşürülmektedir.

“The /r/ sound in the final position of the word *bir* ‘one, certain’ and in the morpheme {-Iyor} is generally dropped in a rapid conversation.”

bir → *bi*

geliyor → *geliyo*

{-Iyor} biçimbiriminde görülen /r/ - düşmesi, {-Iyor} ekinden sonra gelen kişi ekinin ikinci tekil ya da ikinci çoğul kişi eki olduğu durumlarda da görülmektedir:

“/r/ dropping in the morpheme {-Iyor} is also attested when followed by the second singular or plural ending:”

geliyorsun → *geliyosun*

geliyorsunuz → *geliyosunuz* (Özsoy, 2004: 109)

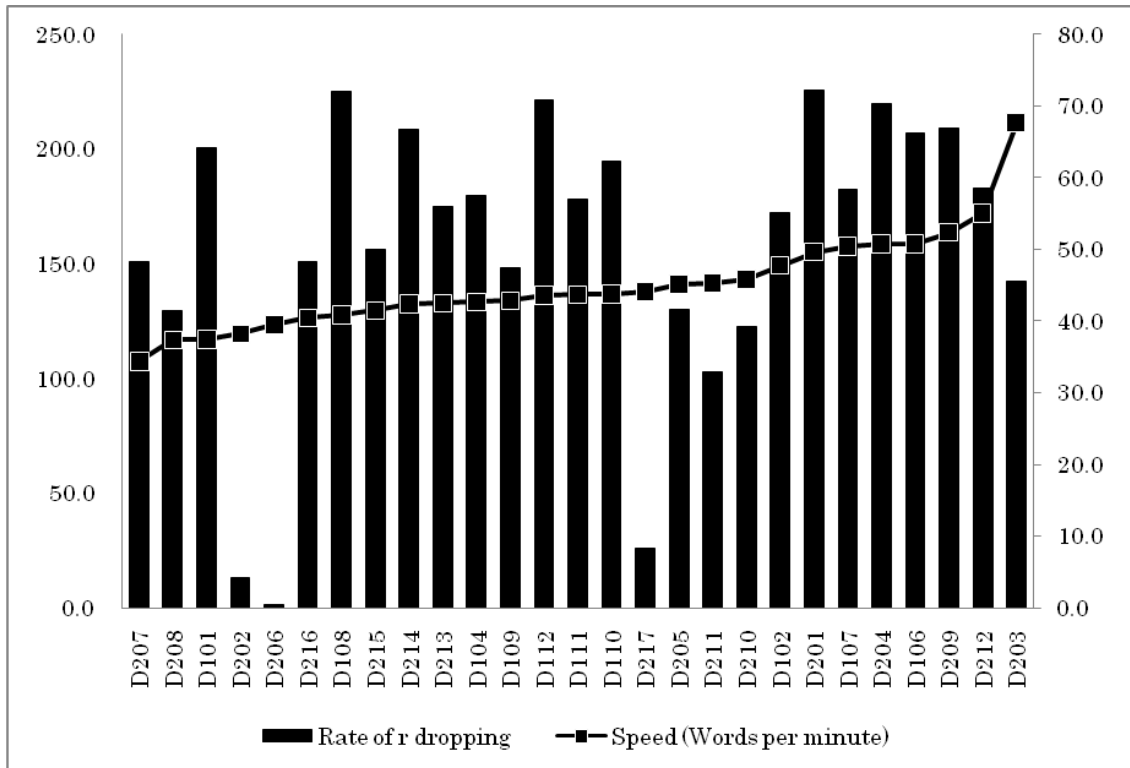
The existence of *-r* dropping has already been recognized in Turkish Linguistics, and many researchers have claimed that *-r* dropping is attested most frequently in rapid speech. They seem to maintain that the final consonant is dropped under the influence of a faster speech style?an accelerated speed of conversation. In fact, there is some degree of truth in this explanation, for the phenomenon of segmental dropping often occurs when speakers are obliged to skip segments or fragments of a word under a certain mental pressure while talking at an unusually high speed. Furthermore *-r* dropping is too well established a phenomenon in spoken Turkish to cast doubt on the above mentioned interpretation. To my knowledge, no serious attempt has been made to demonstratively investigate the phenomenon of *-r* dropping. The main objective of this paper is to redefine the factors of *-r* dropping based on corpus evidence.

3. Factors of *-r* dropping of *-Iyor*

Based on Özsoy's description, we can distinguish two different contexts for *-r* dropping: *-r* dropping in the present progressive suffix *-Iyor* and that in the word *bir*. We will begin with the examination of the different factors of *-r* dropping in the present progressive suffix *-Iyor*.

3.1. Speed and *-r* dropping

Figure 2. Speed and *-r* dropping



First, it seems relevant for us to assess if *-r* dropping can be triggered by a faster speech style. In other words, can rapid speech really provide a favorable condition for *-r* dropping in the present progressive suffix? Speed is calculated here as number of words per minute. In Figure 2, the bar graph represents the percentage of *-r* dropping on the vertical left scale. It is calculated as the number of occurrences of *-Iyo(r)* divided by the sum of the number of occurrences of *-Iyor* and *-Iyo(r)*. Further, the line graph shows the speed of each dialogue. The data are sorted by their speed rate, starting from the slowest D207 on the left to the fastest D203 on the right. Pearson's correlation coefficient is very low: 0.25. This weak positive correlation coefficient $r=0.25$ ($n=27$) is not statistically significant ($p=0.22 > 0.05$). The speed factor is not responsible for *-r* dropping in the present progressive suffix *-Iyor*.

3.2. Topic and *-r* dropping

In everyday conversations, it is often the case that speakers and hearers do not draw attention to some unimportant topics of conversation. Intuition tells us that such an inattentive attitude might trigger some segmental droppings. Different topics of the MSC Turkish are divided into general topics and specific ones. General topics include students, ‘exams, tests’, hometown, dramas, tourist spots, Istanbul, leisure, music, cigarette, and vacation; whereas Turkish culture, spoken language, education, EU and Turkey, and Turkish literature can be considered as specific topics.

Figure 3. -r dropping in general topics

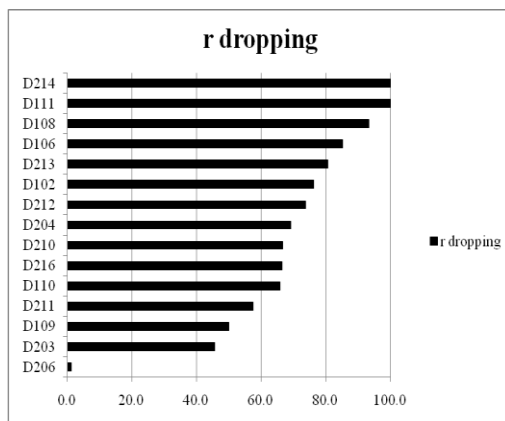
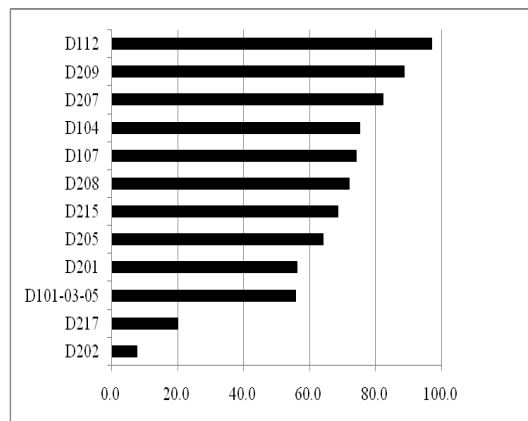


Figure 4. -r dropping in specific topics



Figures 3 and 4 demonstrate that different kinds of topics are not relevant to -r dropping. For instance, our informants talk about ‘leisure’ in two corpora: D108 and D210. The rate of -r dropping is 93% in the former and 67% in the latter. The topic of ‘education’ is always the same in D112 and D201, while their rates of -r dropping are totally different. Pearson’s *r* is very low and the correlation coefficient $r=0.11$ ($n=27$) is not statistically significant ($p=0.60 > 0.05$).

3.3. Grammatical context and -r dropping

Our corpus-driven analysis indicates that -r dropping of *-Iyor* occurs in four different grammatical contexts: definite past, indefinite past, present progressive, and conditional. However, the rate of -r dropping is mostly the same in these contexts; hence, we believe that grammatical contexts are irrelevant to -r dropping, see Figure 5.

Figure 5. -r dropping in grammatical contexts

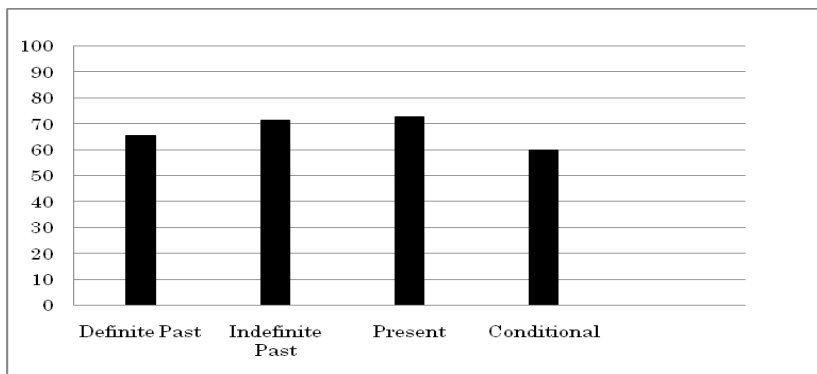


Table 5. Occurrences of -r dropping in grammatical contexts

Definite past	yordu	53	Present progressive	yor	1308
	yorduk	15		yorsun	279
	yordum	43		yorsunuz	91
	yordun	8		yorlar	225
	yordunuz	3		yormu	2
	yorlardı	9		yorki	1
indefinite past	yorlarmış	10	conditional	yorsa	23
	yormuş	32		yorsam	5
	yormuşun	1		yorsan	18
	yormuşum	2		yorsanız	2

In addition, it should be noted that the phenomenon of -r dropping appears mostly in three forms of the present progressive: -yor, -yorsun, -yorlar; these three contexts constitute 85% of all the occurrences of -r dropping, see Table 5.

3.4. Phonetic context and -r dropping

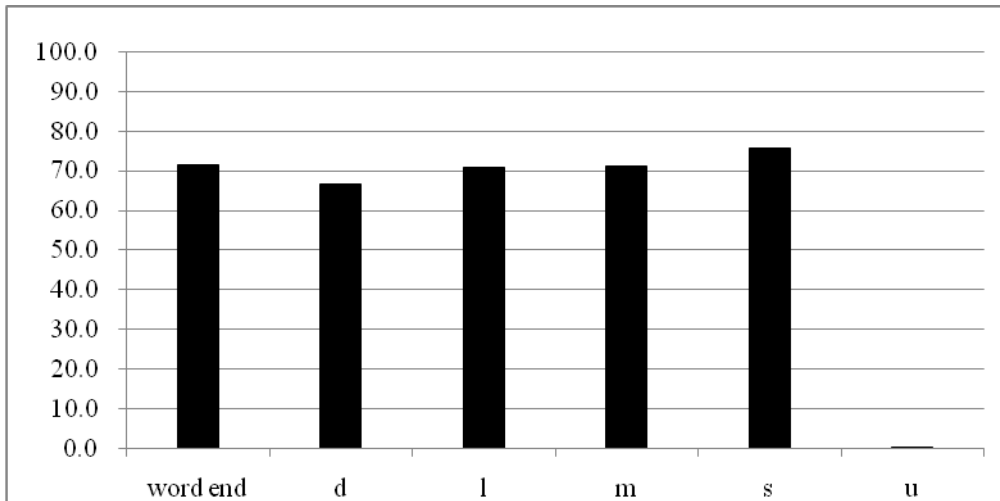
Previous studies tell us that the most influential factor of -r dropping is a phonetic one. For instance, Göksel Aslı and Celia Kerslake describe phonetic contexts of -r dropping in their *Turkish: A Comprehensive Grammar*.

In colloquial speech, the “r” at the end of the suffix is often not pronounced when it is followed by a suffix beginning with a consonant: *gidiyorsun* “s/he goes/ is going” spaces to the right? [ɟ idijosun] / [ɟ idijorsun], (...) or when it appears at the end of a word: *sarıyor* “s/he wraps” [sarijo] /

[sar]jor]space before bracket?. ‘r’ is retained when it is followed by a suffix beginning with a vowel: *seviyorum* “I love/like” [sev]jorum]. (Göksel and Kerslake, 2005, p.83)

Three contexts should be distinguished: 1) *-r* dropping is observed when it is followed by a consonant, 2) *-r* dropping occurs at word end, and 3) there is no *-r* dropping following a vowel. Moreover, the corpus-driven analysis of MSC Turkish will show these phonetic environments in more detail.

Figure 6. *-r* dropping in phonetic contexts



Four consonants that follow the suffix *-Iyor* are /d/, /l/, /m/ and /s/. The rate of *-r* dropping is almost the same as that of word end, which is around 70%; see Figure 6. In general, *-r* is not dropped before the vowel /u/, but four exceptional cases are observed; see Table 6.⁶

Table 6. Occurrences of *-r* dropping in phonetic contexts

context	form	-r dropping	no dropping	Context	form	-r dropping	no dropping
/d/	yordu	53	23	/m/	yormuşun	1	1
/d/	yorduk	15	14	/m/	yormuşum	2	2
/d/	yordum	43	15	/s/	yorsa	23	24
/d/	yordun	8	8	/s/	yorsam	5	5
/d/	yordunuz	3	1	/s/	yorsan	18	1
/d/	Yordur	6	3	/s/	yorsanız	2	2
/l/	Yorlar	225	88	/s/	yorsun	279	79

⁶ However, this exceptional retention of final *-r* may be due to errors in transcription.

/l/	yorlardı	9	8	/s/	yorsunuz	91	23
/l/	yorlarmış	10	4	/u/	yorum	4	921
/m/	Yormu	2	1	/u/	yoruz	0	235
/m/	yormuş	32	11	word end	yor	1308	523

3.5. Sex and age factors in *-r* dropping

Can differences in sex be responsible for *-r* dropping? Figures 7 and 8 show higher scores of *-r* dropping for women. Four women, namely, AK, BH, DH, and NS, drop *-r* in more than 90 per cent of the cases. However, statistically speaking, the correlation coefficient $R=0.19$ ($N=17$) between sex and *-r* dropping is not significant ($p=0.47 > 0.05$).

Figure 7. *-r* dropping in women

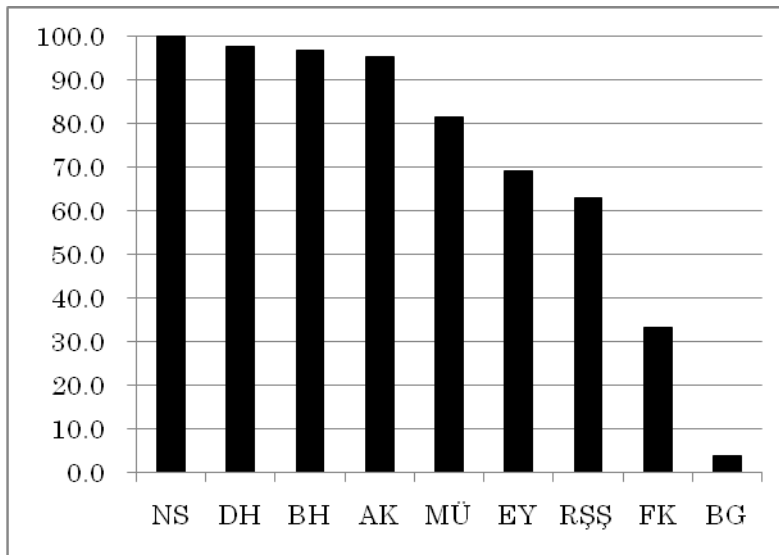
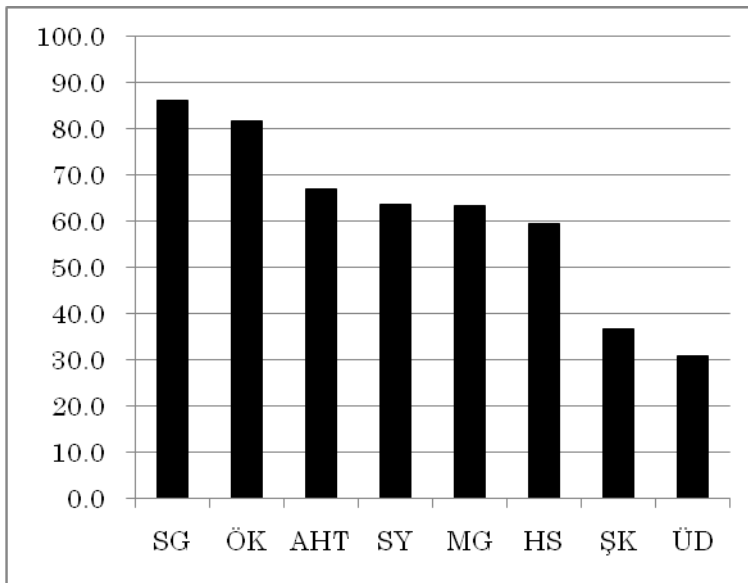
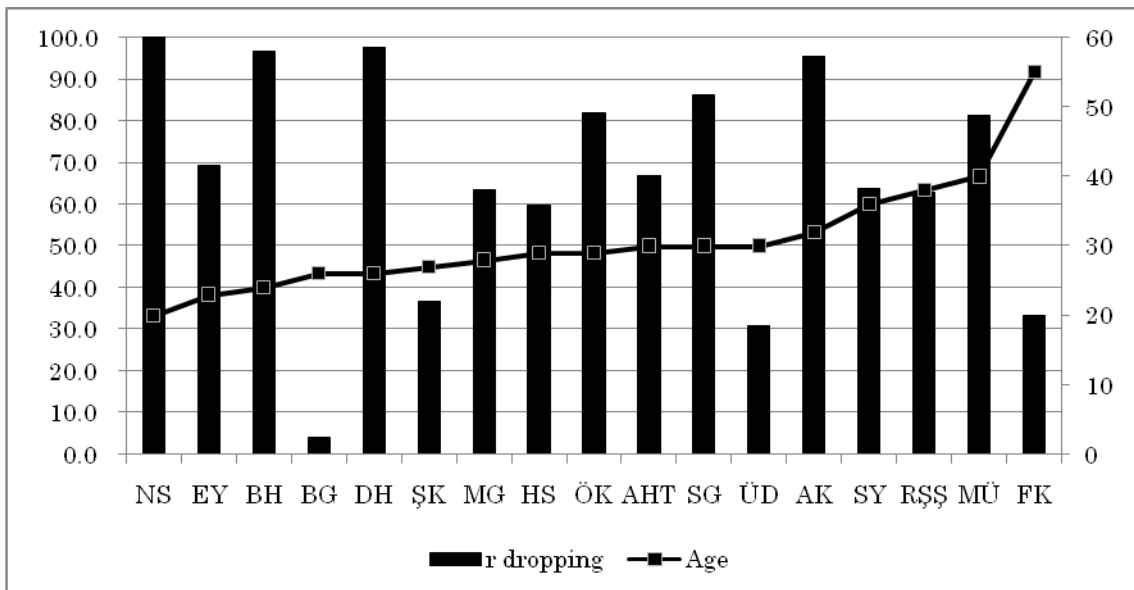


Figure 8. *-r* dropping in men



In Figure 9, the bar graph indicates the percentage of *-r* dropping on the vertical left scale and the line graph represents the age of our informants. The data are sorted by their age, starting from the oldest informant, BH, to the youngest one, NS. Sorry, I don't understand the sentence to the right: Three young informants DH, BH, and NS may indicate their characteristic high rate of *-r* dropping. However, the correlation coefficient $r = -0.28$ ($n = 17$) is not significant ($p = 0.28 > 0.05$).

Figure 9. Age and *-r* dropping



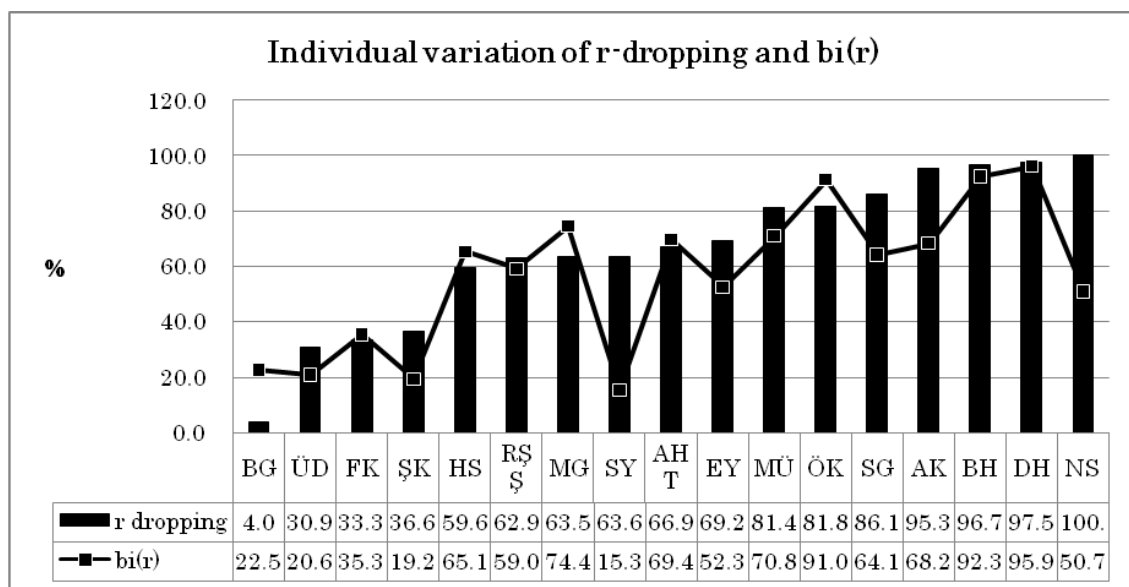
4. Correlation between *-Iyor* and *bir*

In previous sections, we have seen that phonetic conditions are responsible for *-r* dropping. In fact, *-r* dropping can occur before the consonants /d/, /l/, /m/, /s/ and at the word end position, whereas factors of speed, topic, sex, age, and grammatical context are all statistically insignificant for *-r* dropping. However, we still need to interpret how *-r* dropping will be triggered when the abovementioned phonetic conditions are present.

4.1. Individual variation in *-r* dropping of *-Iyor* and *bir*

In Figure 10, the bar graph shows the percentage of *-r* dropping of *-Iyor* on the vertical left scale, while the line graph is that of the word *bir*. The data are sorted by the former in increasing order from BG to NS. The average score of *-r* dropping is 69.0% for *-Iyor* and 61.8% for *bir*. As shown in Figure 10, four informants, i.e., BG, ŞK, SY, and NS, have different phonetic attitudes towards *-Iyor* and *bir*. BG drops the *-r* sound more frequently in *bir* while for ŞK, SY and NS, *-r* dropping is attested more often in *-Iyor*. They are from Izmir, Istanbul, Sivas, and Ardahan, respectively. Their origin cannot explain the reason why they maintain the *-r* sound. Consequently, *-r* dropping seems to be an idiosyncrasy of these informants. In fact, if we exclude the data of these four informants, *-r* dropping in the case of *-Iyor* is strongly correlated with that of *bir*. The correlation coefficient $r=0.85$ ($n=13$) is significant ($p=0.00 < 0.05$). The Null hypothesis is rejected at the threshold of 1%. Therefore, generally speaking, highly educated Turks have the same phonetic attitudes towards *-r* dropping with respect to *-Iyor* and *bir*.

Figure 10. Individual variation of *-r* dropping in *-Iyor* and *bir*



Our hypothesis that *-r* dropping is idiosyncratically motivated is consolidated by the fact that the interpersonal factor may be correlated with *-r* dropping for some informants. Five of our 17 informants show some different interpersonal settings in

their dialogues. In one situation which is supposed to be more or less relaxing and comfortable, he or she has a conversation with a partner of almost his or her age. In the other situation, he or she chats with a partner who is ten years older than him or her. Our informants' behaviors are sharply distinguished into two different types.

For the informants who maintain the same phonetic attitudes toward both *-r* droppings in *-Iyor* and *bir*, the shift from *-r* dropping to non *-r* dropping is generally excluded. They retain mostly the same proportion of *-r* dropping even though their partner is of a similar age or about ten years older. For instance, *-r* dropping and non *-r* dropping rates are not very different for HS, who is 29 years old. His partner in D201 and D204 is RŞŞ, who is 38 years old. On the contrary, in D212, HS is talking with ÖK, who is of the same age. The same rule holds good for RŞŞ and ÖK. These two informants do not change their *-r* dropping rate in accordance with their partners; see Table 7.

Table 7. Interpersonal factor: individuals making no shift

	<i>-r</i> dropping	non <i>-r</i> dropping	<i>-r</i> dropping	non <i>-r</i> dropping
	D201, D204		D212	
HS	66.0%	34.0%	57.3%	42.7%
	D213		D201, D203, D204, D205	
RŞŞ	54.8%	45.2%	57.1%	42.9%
	D209, D212		D203	
ÖK	72.5%	27.5%	65.3%	34.7%

ÜD and ŞK, who demonstrated a very low proportion of *-r* dropping, show a keen shift of *-r* dropping in accordance with their partners. In D202, 30 years old ÜD has a conversation with BG who is 26 years old and rarely drops the *-r* sound, whereas in D205 and D207, the *-r* dropping rate increases to 32.7% in his dialogues with partners of 38 and 40 years old respectively. In D206, D210, and D211, the 27-year-old ŞK is talking with partners who are 26 and 30 years old, and *-r* dropping rarely occurs. The rate goes suddenly up to 30% when he talks to his 40-year-old partner; see Table 8. It does seem to us that the shift from non *-r* dropping to *-r* dropping is an idiosyncratic phenomenon, though our corpus data are undoubtedly insufficient to fully substantiate this claim.

Table 8. Interpersonal factor: individuals with shift

	<i>-r</i> dropping	non <i>-r</i> dropping	<i>-r</i> dropping	non <i>-r</i> dropping
	D205, D207		D202	
ÜD	32.7%	67.3%	6.8%	93.2%
	D208		D206, D210, D211	
ŞK	30.1%	69.9%	13.1%	86.9%

4.2. Factors of *-r* dropping in the word “bir”

Finally, we will conduct the same statistical analysis of *-r* dropping in the word *bir*. According to Table 9, topic, sex, and age factors are all irrelevant to *-r* dropping, whereas speed and *-r* dropping are correlated. The correlation coefficient $r=0.46$ ($n=27$) is significant ($p=0.02 < 0.05$).

Table 9. Factors of *-r* dropping in *bir*

topic	$r = -0.17$ ($n=27$), $p = 0.41 > 0.05$	NS
sex	$r = 0.16$ ($n=17$), $p = 0.53 > 0.05$	NS
age	$r = -0.21$ ($n=17$), $p = 0.42 > 0.05$	NS

Five informants, namely, BG, FK, SY, ŞK, and ÜD, have a preference for retaining the *-r* sound, see Figure 9.

As for phonetic context, a high frequency of *-r* dropping is attested, particularly before the four consonants, /ş/, /d/, /k/ and /y⁷/; see Figure 11.

Figure 11. Phonetic contexts of the occurrences of *bir*

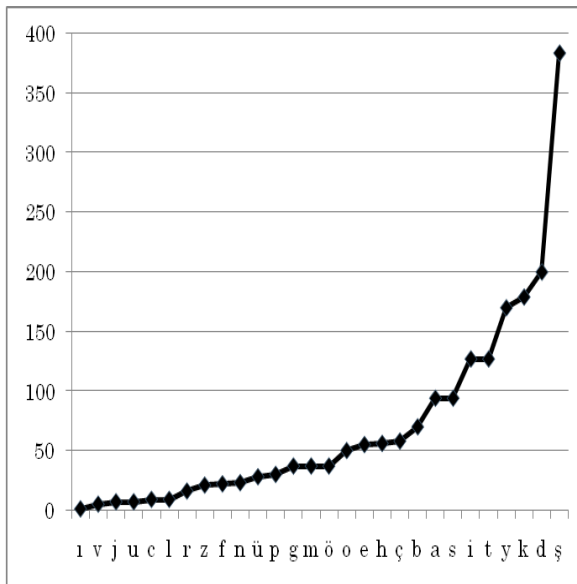
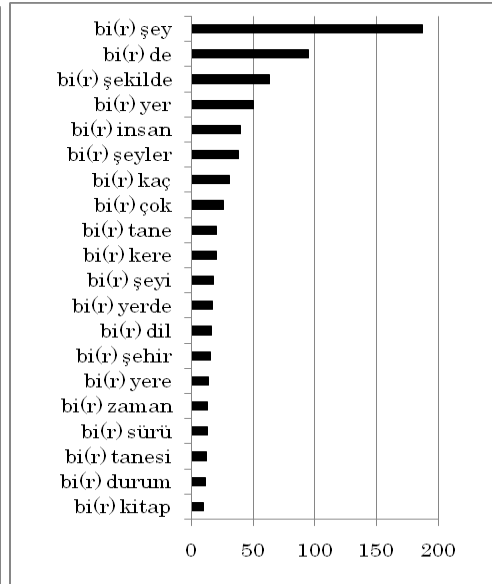


Figure 12. Collocations of *bir*



However, this apparent high frequency is due to the collocations of the word *bir*, for e.g., *bir şey*, *bir de*, *bir şekil*, and *bir yer*. This small word often appears along with *şey* (thing), *de* (and), *şekil* (form), and *yer* (place); see Figure 12.

Conclusion

⁷ The notation of /y/ comes from Turkish alphabetical transcription and represents the semi-vowel /j/.

As already claimed in previous studies, two phonetic contexts should be taken into consideration for the analysis of *-r* dropping in the present progressive suffix *-Iyor*. They are “word end” and “before consonants.” The present corpus-driven analysis specifies four consonants /d/, /l/, /m/, and /s/. For the word *bir*, the collocations *bir şey*, *bir de*, *bir şekil*, and *bir yer* can be linked to a high frequency of *-r* dropping.

What are the factors that have the highest influence on *-r* dropping? We investigated whether sociolinguistic factors such as sex, age, and birth place were relevant to *-r* dropping. These sociolinguistic factors are all statistically insignificant. Other stylistic factors such as the topic of conversation and speed were also examined. Only speed is statistically significant in *-r* dropping in the case of *bir*, while the topic of conversation never leads to this phenomenon.

Our corpus-driven analysis demonstrates that *-r* dropping is not sensitive to sociolinguistic factors but individually motivated. In fact, some informants exhibit a tendency to avoid *-r* dropping in *-Iyor* or *bir*, but at the same time, other informants have exactly the same tendency of *-r* dropping in both *-Iyor* and *bir* without any sociolinguistic difference based on sex and age. Interestingly enough, similar to our *-r* dropping in Turkish, Dufter and Stark (2007) have claimed that the dropping of the negative particle *ne* in French is independent of the age, sex, and socioprofessional status of speakers⁸. Further, citing the dichotomy of universal traits versus individual traits in communicative moments, they stated that individual language traits are contingent and idiosyncratic.⁹

Finally, it is important to note that the present corpus-driven analysis has some limitations. First, our corpus is too small to provide a global view of *-r* dropping in spoken Turkish. Second, the 17 informants, who are all highly educated Turks, represent only a very small portion of the complex Turkish language community. We must therefore be careful not to draw hasty conclusions. Nevertheless, it is our conviction that a large scale corpus-driven analysis is necessary if we want to go beyond traditional descriptions and assumptions.

References

- Aksan, D. (ed). 1978. *Türkiye Türkçesi Gelişmeli Sesbilimi*. Ankara: Türk Dil Kurumu Yayınları.
- Delais-Roussaire, É. and J. Durand (eds). 2003. *Corpus et variation en phonologie du français*. Toulouse: Presses Universitaires du Mirail.
- Demircan, Ö. 2001. *Türkçenin ses dizimi*. İstanbul: Der Yayınları.
- Dufter, A. and E. Stark. 2007. “La linguistique variationnelle et les changements linguistiques ‘mal compris’: le cas du *ne* de négation”. *Études sur le changement linguistique en français*, B. Combettes and Ch. Marchello-Nizia (eds). Nancy: Presses Universitaires de Nancy. 0000-0000.

⁸ Dufter and Stark (2007) p.116.

⁹ Cf. Koch and Oesterreicher (2001).

To appear in *Corpus Analysis and Variation in Linguistics*, Yuji Kawaguchi, Makoto Minegishi, Jacques Durand (eds), series TUFSL (Tokyo University of Foreign Studies : Studies in Linguistics), vol.1, Amsterdam/Philadelphia, John Benjamins, 2009.

Gibbon, D., R. Moore and R. Winski (eds). 1998. *Handbook of Standards and Resources for Spoken Language Systems* Vol.1–4. Berlin/New York: Mouton de Gruyter.

Göksel, A. and C. Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.

Kawaguchi, Y. 2005. “Phonological System of Modern Turkish – A critical review on phonemes”. *Journal of the Institute of Language Research*10, The Institute of Language Research, Tokyo University of Foreign Studies. 35–62.

Kawaguchi, Y. 2006. “Usage-Based Approach to Linguistic Variation – Evidence from French and Turkish”. *Spoken Language Corpus and Linguistic Informatics*, Y. Kawaguchi, S. Zaima and T. Takagaki (eds). Amsterdam/Philadelphia: John Benjamins. 247–267.

Koch, P. and W. Oesterreicher. 2001. “Gesprochene Sprache und geschriebene Sprache”. *Lexikon der Romanistischen Linguistik*, Vol.1.2 . G. Holtus, M. Metzeltin, C. Schmitt (eds.). Tübingen: Max Niemeyer.584–627.

Kornfilt, J. 1997. *Turkish*. London: Routledge.

Morel, M.-A. and L. Danon-Boileau. 1998. *Grammaire de l’intonation. L’exemple du français*. Paris: Ophrys.

Özsoy, A. S. 1999. *Türkçe Turkish*. İstanbul: Boğaziçi Üniversitesi Yayınları.

Özsoy, A. S. 2004. *Türkçenin yapısı – I*. İstanbul: Boğaziçi Üniversitesi Yayınları.

Sarıca, M. 2005. *Yeni dilbilim kuramlar ışığında Sözlü Dil Yapısı*. İstanbul: Multilingual.

Selen, N. 1979. *Söyleyiş sesbilimi akustik sesbilim ve Türkiye Türkçesi*. Ankara: Türk Dil Kurumu Yayınları.

Teubert, W. and A. Cermáková. 2004. *Corpus Linguistics A Short Introduction*. PLACE: Continuum.

Yılmaz, S. 2006. “Présentation d’un corpus oral en ture contemporain”. *Aspects of Corpus Linguistics – Spoken Language Corpora and Written Language Corpora-*, Working Papers of Linguistic Informatics, 12. Tsuruga & al., 21st Century COE “Usage-Based Linguistic Informatics,” Graduate School of Tokyo University of Foreign Studies, Forename, Name (ed) 201–240.

AntConc 3.2.1w (Windows). Laurence Anthony. Faculty of Science and Engineering Waseda University, Japan; anthony@waseda.jp.