

Development of lexical richness in Turkish written texts of bilingual adolescents in Germany

Carol W. Pfaff* – Seda Yılmaz** – Meral Dollnick** – Mehmet-Ali Akıncı***

1. Introduction

This paper explores quantitative and qualitative aspects of the development of lexical richness in Turkish from our ongoing study of the language development in their first, second and third languages of children and adolescents with migrant background from Turkey living in Germany and France. Here we focus on the development of the lexicon in written texts in Turkish in a longitudinal subsample of pupils in 10th and 12th grade pupils in Berlin.

One component of lexical richness is lexical diversity, often quantified as type-token ratio (TTR) or related measures such as the index of Guiraud. However, as pointed out by Daller et al., who focus on measures sensitive to the proportion of advanced vocabulary in oral production, it is essential to go beyond quantitative measures to assess qualitative aspects of the lexicon as well. In this paper we pursue the linguistic features with underlie judgments of “advanced” vocabulary. Additionally, we explore the possible correlations of lexical richness with individual sociobiographic factors and pupils’ language use outside the classroom, in particular the extent of their formal instruction in Turkish.

In our approach to the study of the lexical richness here we draw on data from our ongoing bi-national study of the language development in their first, second and third languages of children and adolescents with migrant background from Turkey living in Germany and France (Akıncı & Pfaff 2008, Pfaff 2009, Pfaff, Schroeder & Dollnick 2009, Akıncı, Pfaff & Dollnick 2010). These data are based on oral and written texts collected in response to a short video without dialogue developed for the cross linguistic study of later language development in monolinguals (Berman & Verhoeven 2002). In the present paper, we examine the development of lexical diversity of the written texts of a subset of pupils who participated both as 10th and 12th graders.

It is clear that not all children and adolescents who live in the same area and attend the same school have similar sociolinguistic patterns. Individual differences in vocabulary are related to the participants’ and their families’ sociobiographies and language use patterns (Akıncı & Pfaff 2008). Our previous papers have shown differences in syntactic development (Pfaff, Akıncı & Dollnick 2009), orthography and written usage (Pfaff, Schoeder & Dollnick 2009, (Akıncı, Pfaff & Dollnick 2010). In the present paper we explore individual differences in lexical development and lexical richness evidenced in their texts. One factor, we expected to be particularly likely to play a role is the extent to which the participants have had formal instruction in Turkish.

* Freie Universität Berlin.

** Universität Potsdam.

*** CNRS & Université Lyon.

2. The LLDM / MULTILIT longitudinal subsample

In the present paper we focus on the written texts of a longitudinal subsample consisting of 11 secondary school pupils at a gymnasium in Berlin-Kreuzberg who participated in the LLDM study as 10th graders in 2008 and in the MULTILIT study as 12th graders in 2010.¹ According to their self-reports on participation in Turkish classes, this subsample is almost equally distributed in three subgroups, as shown in the rightmost column of Table 1: Those with “0” did not participate in formal Turkish classes but had contact with written Turkish media and informal instruction in their families or with other relatives; those with “1” participated in Turkish classes at the primary level, either in school as part of a bilingual literacy program or outside school; those with “2” were participating in Turkish classes in their secondary school, where Turkish was an option as their second foreign language. The details on the participants are shown in Table 1:

Table 1. Longitudinal subsample: Berlin -10th and 12th grades

Pseudonym	Sex	First exposure to German	Age: 10th grade	Age: 12th grade	L1	Turkish instruction
Asli	F	School	17;08	19;08	TR	0
Ela	F	Family	16;11	18;11	TR	0
Neslihan	F	School	18;04	18;04	TR	0
Vedat	M	Kindergarten	15;08	17;08	TR	0
Ismail	M	Family	16;02	18;02	TR	1
Kemal	M	Family	16;08	18;08	TR	1
Serhat	M	Family	16;02	18;02	TR	1
Ayhan	M	Family	17;07	19;07	TR	2
Hacer	F	Family	16;01	18;01	TR	2
Nihal	F	Family	16;05	18;05	TR	2
Yaşar	M	Family	16;04	18;04	TR/KU	2

0=no formal instruction, 1=instruction at primary school age, 2 = instruction in secondary school

1 Since 2007, we have been engaged in a comparative study of language development in pupils with migration background from Turkey in Germany and France, eliciting oral and written texts about personal experiences and (Pfaff 2009, Akıncı et al. 2010) and self-reported data on background and language practices with family and friends outside of school. (Akıncı & Pfaff 2008). We are indebted to the DAAD-PHC for funding the Later Language Development (LLDM) project, led by Pfaff and Akıncı 2007-2009 and to the DFG-ANR for funding the project as the Multiliteracy (MULTILIT) project, led by Schroeder and Akıncı, in cooperation with Pfaff 2010-2012.

3. Research questions and preliminary hypotheses on development of lexical richness

We hypothesize that text length, lexical diversity and lexical richness are influenced by several factors, including the following investigated here:

- Older pupils will produce longer texts, more diverse, more advanced vocabulary
- Expository texts will be longer and have more diverse and more advanced vocabulary than narrative texts
- Pupils with more instruction will produce longer texts, with more diverse and more advanced vocabulary

4. Analysis and results

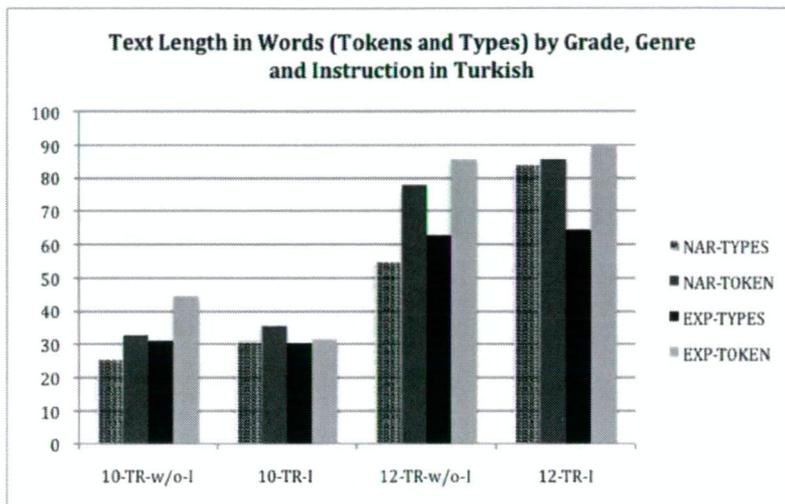
4.1. Quantitative analysis

Turning now to the quantitative analysis of all 11 individuals in the subsample, we find the following results for text length, based on a simple word count (excluding words crossed out by the participant) and lexical diversity, based on the type/ token ratio. As noted by Strömquist et al. (2002: 60) and also by Daller et al (2003: 197), it is essential to exercise great care in making quantitative comparisons of typologically unrelated languages. While this paper is concerned only with the Turkish, our study encompasses the parallel investigation of the other languages in the participants' verbal repertoires, German, French and English which differ typologically from each other and especially widely from Turkish in their lexical properties. For our calculation of types in Turkish in the present paper, we adopted the following conventions:

- Word forms differing only in inflectional morphology e.g., *okul*, *okulda*, count as tokens of the same type
- Word forms differing in derivational morphology e.g., *Alman*, *Almanca*, count as tokens of different types
- Complex verbs with multiple words e.g., *kavga etmek*, *kopya çekmek*, count as 1 lexeme type
- Complex adverbials with multiple words e.g., *çoğu zaman* or *o zaman* count as two types

Figure 1 shows the results for individuals, by grade, genre and participation in Turkish classes. At both grades the texts produced by the pupils were very short, ranging from 17 to 157 words in length. The 10th grade texts ranged from 17–75 words and the 12th grade texts ranged from 50–157 words.

Figure 1. Text length in words (tokens and types) by grade, genre and instruction in Turkish



Legend: TR-w/o-I (Group 0), TR-I (Groups 1 and 2)

As shown in Figure 1, pupils produced notably longer texts at 12th grade as expected. This held for each individual as well as for the groups. Contrary to our expectations, we found relatively little repetition of the same types. This may be an artifact of the elicitation procedure which required only very short texts and allowed the participant to choose whether to discuss just one incident or several topics.

There was no consistent clear relationship between genre and text length or lexical diversity. At the 10th grade, 7 of the 11 pupils had longer personal narrative texts but at 12th grade more pupils (again, 7 of the 11 but not the same individuals) had longer expository opinion texts. We speculate that this may be due the increasing emphasis on argumentative text production in school, especially since there was a relationship text length and genre to formal instruction in Turkish: at the 10th grade there was no consistent relationship to formal instruction but at the 12th grade, those with formal instruction in secondary school produced longer texts.

4.2. Qualitative analysis: advanced vocabulary

As also noted by Daller et al. 2003, the classification of “advanced” vs. “basic” vocabulary presents non-trivial problems. Their solution was to supplement the published Tezcan 1988 word list by judgments of teachers of Turkish as a foreign language in Turkey. We similarly included native speaker judgments but, in our case, the judges were graduate students and Turkish teachers in Berlin who were familiar with Turkish both in Berlin

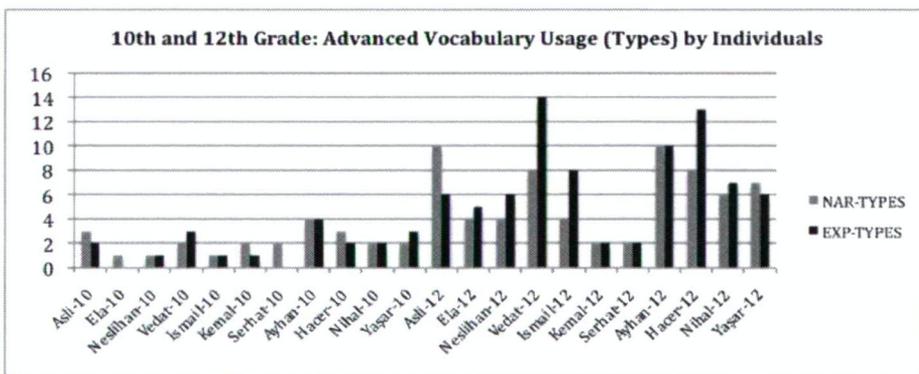
and in Turkey. We considered several different aspects of “advanced” vocabulary: morphological complexity, etymology of the root, abstract vs. concrete, degree of specificity of the lexical items, searching for an appropriate categorization relative to the age, language dominance and proficiency of the participants in the study and sensitive to the topic and situational context as well.

Thus, words such as *genellikle* ‘generally’ which are morphologically complex but relatively common formulaic expressions, may, in context be classified as “basic” rather than “advanced” as a result of their high degree of entrenchment in the verbal repertoires of older children, adolescents and adults. The same may apply to the classification of “abstract” words or collocations which are “entrenched” i.e., “common” or “frequent” in the contexts the participants discussing, such as expressions for the activities shown in the film such as *kavga etmek* ‘fighting’, *kopya çekmek* ‘cheating’ or nominalizations such as *dışlanma* ‘discrimination or ostracism’, which are frequently discussed topics.

Another aspect of advanced vocabulary, also mentioned by Daller et al. is the correlation of qualitative and quantitative measures of vocabulary with syntactic complexity, measured in terms of complex embedded clauses with *-mA*, *-(y)An*, *-DIK*, less frequent, more difficult constructions as than are *-mAK*, gerunds and postposed participles. We also find that complex syntax and complex vocabulary go together, as shown in the examples cited in Pfaff et al. 2009, to be further elaborated in a future publication.

For the present paper, we classify “advanced” vocabulary items by relying on the judgments of native speakers of Turkish who are familiar with the linguistic development of children and adolescents raised in communities in Northwestern Europe, where the varieties of Turkish heard and used by the participants in their daily lives differ from those generally taught to foreigners, reporting here on the judgments of the co-authors of the present paper, thus differing crucially from the teachers of Turkish as a foreign language in Turkey consulted by Daller et al. 2003.² Our results are shown in Figure 2:

Figure 2. Advanced vocabulary: types by grade and genre by individual, 10th and 12th grades



² Subsequently the word lists for the oral as well as the written texts, adjusted to standard orthography, were presented to a panel of teachers of Turkish as a heritage language in Berlin for their judgments. The results will be reported in a later publication.

As shown in Figure 2, advanced vocabulary increases with age, as expected. However, the predicted effect of instruction in Turkish was not found. At 10th grade, the expectation that formal instruction would have a positive effect on use of advanced vocabulary appears to be confirmed. At 12th grade, those with instruction at secondary school, considerably increased their usage of advanced vocabulary, as expected. However so did some of the pupils with no formal instruction, particularly Vedat, perhaps as a result of individual motivation and to contact with classmates who do take Turkish as second foreign language. This aspect deserves further investigation with attention to their self reports and, if possible, with follow-up group interviews to be conducted in the following months, their last year of secondary school.

Some examples are shown in (1–8) below.

Advanced (or academic) vocabulary:

(1) *milliyet* ‘nationality’ SERHAT-TI-1-WN-10, (2) *ahlak* ‘morals’ VEDAT-TI-0-WE-12

Morphological complexity: nominalized verbs; converbs:

(3) *konuşarak* ‘talking’(adv.) ASLI-TI-0-WE-10, (4) *görülmesi* ‘it’s been seen’ (passive) SERHAT-TI-1-WN-10

(5) *düşürttüğü* that s/he made something fall ‘dropped’ ASLI-TI-0-WE-12, (6) *dışlamaması* ‘his/her not excluding (someone)’ ASLI-TI-0-WE-12

Text-structuring expression:

(7) *bence* ‘I think’ ASLI-TI-0-WE-10, (8) *demek istediğim* ‘what I wanted to say’ İSMAİL-TI-1-WE-10

Our initial hypothesis that “advanced vocabulary” would be found more in expository than in narrative texts was not confirmed in the present study; we found no consistent relation to genre in this sample. Although most pupils have more advanced vocabulary in expository texts, some have more in narrative texts.

The effect of spoken vs. written modality was not addressed in the present paper, which is limited to written work. We find considerable amounts nominalized forms used here, but comparison with the spoken texts of these participants is left to a future paper.

4.2.1. Use of idiomatic expressions

The use of idiomatic expressions is indication of the participants’ lexical / phraseological development through the actual use of Turkish.

(9) *Canı geleceğine mala gelsin derler.*

‘It is said, better that your material possessions are harmed than your life’ YAŞAR-TI-2-WE-12

Participants who have not had formal instruction in Turkish also make use of idiomatic expression, though these are not always completely native-like, as in (10)–(11) from Vedat:

(10) ...*selam veririm, gönlünü hat(ı)rını sorarım, sonra yine istersem arkadaşlarımın yanlarına giderim.* VEDAT-TI-0-WE-10

In this example Vedat combines parts of two idiomatic expressions: (*gönül almak* 'to take heart' *hatır sormak* 'ask somebody how he is')

(11) *Arkadaşını göster söyleyeyim sen kimsin diye.*

'Show me who your friends are and I'll tell you who you are'

VEDAT-TI-0-WE-12

A particularly interesting example is found in Vedat's expository text at the 12th grade in (12):

(12) *İlam gel tertip, hasbahim ol diyecek değilsin.*

'You are not going to say come on dude, lets make friends' VEDAT-TI-0-WE-12

The intended meaning was not obvious due to the inclusion of idiosyncratic lexical items which were unknown to us. Here we follow the interpretation suggested by Işıl Erduyan, who suggests that *ilam* may be a misspelling of *illa*, used for emphasis in an imperative act, roughly 'for sure' and that *hasbahim* (root *hasbah*, could be a misinterpretation of *hasbihal* 'an enjoyable conversation'. *Tertip* is a 'good male friend', a term only men use among themselves, stemming from its use as a name given to the males from the same group of army who start military duty at the same time.

4.2.3. Oral features in written texts:

Examination of the word lists revealed that the oral pronunciation forms *bi* occurs in the speech of 3 pupils who either did not have any formal instruction in Turkish or had this only at the primary level.

(13) *Dışlanırsada bi kendisini yoklaması gerek.* VEDAT-TI-0-WE-12

(14) *Bi insan bana nasıl davranırsa bende ona aynı şekilde bende ona öyle davranırım.* KEMAL-TI-1-WE-10

The form *bi* does not occur in the written texts of those with instruction in Turkish at the secondary level.

(15) *Almanlar bizim yanımıza geldiki zaman bazı arkadaşlarımız kalkiyo gidir yada ki hiç konuşmuyorlar.* KEMAL-TI-1-WN-12

4.2.4. Cross-linguistic influence from German: loanwords, calques and orthography

Loan words from German do not occur in the written Turkish texts examined here, though a few were found in the oral texts, as we discuss in Pfaff, Woerfel and Yılmaz 2011.³ In a few cases, we find collocations which may be calques based on German, as in

3 German words occurred more frequently in the English oral and written texts.

example (16) and transfer of German orthographic conventions such as capitalization of nouns and substitution of letters for voiced and voiceless sounds.

(16) *Öğretmen bizi ilk gördüğünde bir kere “okuldaki yasaları – kâğıtini” verdi ve onu herkez bir defa yazmalıydı.* ELA-TI-0-WN-10
 ‘When the teacher saw us the first time, one time she gave us the ‘school law paper’ and everybody should write that once’.

It is noteworthy that Ela put this phrase in quotes, indicating her metalinguistic awareness of the special (if not calqued) nature of the school-related jargon.

Orthographic transfer of orthographic conventions from German stand out in the written texts of several pupils, particular those who have had no formal instruction in Turkish or not since primary school, for example in (17) and (18): //capitalization of nouns as in (17) and sound-spelling conventions for voicing in (18)-(22).

(17) *...önce başka okulda Okulda karkisma tartışmalar oluyordu ve bu Okul atmosfere çok bozuyordu.* SERHAT-T1-1-WN-12
 ‘... earlier in another school, there were discussions at school and this was ruining the atmosphere a lot.

(18) *Diyelim ben birisini arkadaş olarak sevmiyorum, o saman ben o kişinin yanına gidib selam veririm, gönlünü hat(ı)rını sorarım, sonra yine istersem arkadaşlarımın yanlarına giderim.* VEDAT-T1-0-WE-10

(19) *atmosphärede* for *atmosfer*, SERHAT-TI-1-WE-12

(20) *Aggressivlestigini* YAŞAR-TI-2-WN-10, (21) *Aggressivlesirim* YAŞAR-TI-2-WE-10

(22) *Mesela bir insan parayı bulduhunda yada başka bir şeyde olabilir geri vermedidir çünkü o kişi o bulduhu eşyayı kendi gücüyle kazanmamıştır.* KEMAL-TI-1-WE-10

Orthographic *h* in German makes the preceding vowel long, just as *ğ* in Turkish does.

5. Discussion and perspectives for further investigation

This exploratory study raised several questions about the effect of age, genre and formal instruction in Turkish on lexical diversity and the type of lexical items employed. Our findings were that:

AGE / GRADE clearly correlates not only with text length but also with diversity as measured by TTR. The use of advanced / academic / morphologically complex vocabulary is higher at the 12th grade, as expected. In further work reported in Pfaff, Woerfel and Yılmaz 2011, we found that for Turkish (and also for German and English) 10th grade, oral texts showed more lexical diversity than written texts. But at 12th grade, written texts are more lexically diverse. We attribute this to the increasing emphasis on written production in the higher grades.

GENRE was not found to correlate with consistent difference. We speculate that there are two reasons for this finding: first, that the actual nature of the participants' production is mixed in both the "personal narrative" and in the "expository" texts, which, when examined closely turn out to have reported events and evaluative statements in the passages.

TURKISH INSTRUCTION was not found to play as important a role in the usage of advanced or complex lexical items as we had hypothesized. We speculate that that this very probably can be attributed to the participants having learned written Turkish outside of formal classes – with relatives and informally through exposure to written Turkish print and other media in the neighborhood, which is very Turkish-dominant in many settings. We did find a correlation with Turkish instruction and orthography, such that informal short forms like *bi* were written only by those without formal instruction in secondary school, though all participants used such forms orally.

With respect to language contact phenomena, we found no actual loan words from German in the written texts. We did find transfer of some aspects of German orthography and some evidence of possible calques on German collocations, but this will be treated in more detail in a further paper.

Further differentiation of lexical items to refine our notions of basic and advanced vocabulary is necessary. In addition to including further research on lexicography and phraseology, we have conducted an additional survey of Turkish heritage language teachers in Berlin who have an extensive knowledge of the linguistic ecology of and setting of bilingual children such as those investigated in the present paper. These results will be included in a later paper.

Finally, we find it essential to pursue the qualitative dimension of lexical items in context, considering the relative entrenchment of particular items and constructions as they develop in the cognitive representations of individuals. While the cognitive representations are not directly available, the analysis of written texts such as those we have analyzed here and the spoken texts yet to be analyzed will surely provide an appropriate window on this aspect of language development, not only in the heritage language, but in the dominant and foreign languages as well.

References

- Akıncı, M. A. & Pfaff, C. 2008. *Language Choice, Cultural and Literacy practices of Turkish bilingual adolescents in France and in Germany*. International Association for Applied Linguistics (AILA) Essen.
- Akıncı, M. A. & Pfaff, C. W. & Dollnick, M. 2010. Orthographic and morphological aspects of written Turkish in France, Germany and Turkey. In: Ay, S. & Aydın, Ö. & Ergenç, I. et al. (eds.) *Essays on Turkish Linguistics*. Wiesbaden: Harrassowitz. 363–372.
- Berman, R. A. & Verhoeven, L. 2002. Developing text-production abilities across languages, genre and modality. *Written Languages and Literacy* 5: 1, 1–44.
- Daller, H. & van Hout, R. & Treffers-Daller, J. 2003. Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics* 24: 2, 197–222.

- Pfaff, C. W. 2009. Parallel assessment of oral and written text production of multilinguals: Methodological and analytical issues. In: Ahrenholz, B. (ed.) *DaZ-Forschung. Empirische Befunde zum Deutsch-als-Zweitsprache-Erwerb und zur Sprachförderung. Beiträge aus dem 3. Workshop Kinder mit Migrationshintergrund*. Freiburg in Breisgau: Fillibach. 213–233.
- Pfaff, C. W. & Akıncı, M. A. & Dollnick, M. 2009. Complex sentences in the written texts of Turkish trilingual adolescents in Germany and France. Paper presented at the *Seventh International Symposium on Bilingualism (ISB7)* Utrecht.
- Pfaff, C. W. & Akıncı, M. A. & Dollnick, M. 2009. Complex sentences in the written texts of Turkish trilingual adolescents in Germany and France. Paper presented at the *Seventh International Symposium on Bilingualism (ISB7)* July 8–11, 2009. Utrecht.
- Pfaff, C. W. & Schroeder, C. & Dollnick, M. 2009. Türkischer Schriftspracherwerb in der mehrsprachigen deutschen Gesellschaft: Einflüsse aus deutschen und türkischen Varietäten. *XI. Türkischen Internationalen Germanistik Kongress*, Izmir.
- Pfaff, C. W. 2010. Multilingual development in Germany in the crossfire of ideology and politics. In: Okulska, U. & Cap, P. (eds.) *Perspectives in Politics and Discourse*. Amsterdam: Benjamins. 328–357.
- Pfaff, C. & Woerfel, T. & Yılmaz, S. 2011. Development of lexical richness in Turkish, German and English of multilingual adolescents in Berlin. Paper presented at the *International Symposium on Bilingualism (ISB8)*. June 15–18, 2011. Oslo.
- Strömquist, S. & Johansson, V. & Kriz, S. & Ragnarsdóttir, H. & Aisenman, R. & Ravid, D. 2002. Toward a crosslinguistic comparison of lexical quantity in speech and writing. *Written language and literacy* 5: 1, 45–69.
- Tezcan, N. 1988. *Elementarwortschatz Türkisch-Deutsch*. Wiesbaden: Harrassowitz.
- Treffers-Daller, J. & Özsoy, A. S. van Hout, R. 2007. (In)complete acquisition of Turkish among Turkish-German bilinguals in Germany and Turkey: An analysis of complex embeddings in narratives. *International Journal of Bilingual Education and Bilingualism* 10: 3, 248–276.