

A Multilingual Learner Corpus: A Description of the Kenyan Multilingual Individual and His L4¹ Interlanguage

Victoria Anyango Oketch²

Abstract

This study focuses on the cross-linguistic interaction in the written production of the Kenyan multilingual learner of German as a foreign language. Examples are given that show that the previous knowledge of languages other than the mother tongue of the subjects involved plays a role in influencing written lexical production. Thus the German interlanguage under study, which is being learnt as the learners' first foreign language, reveals influence due to interlanguage transfer resulting from the presence of more than two language systems. The paper explores some aspects of the phenomenon of multilingualism, thus a description of the trilingual competence of the subject under study is provided.

1. Introduction

Multilingualism is a common pattern if not the norm in many parts of Africa (Childs 2003: 21). A typical citizen of Kenya will speak at least two or three different languages during the course of a day, using them for different purposes. The focus of this study is to investigate the nature of the learner language of their first foreign language, German, which, according to a chronological classification, would be the L4. The research study attempts to answer the following question: Which role does English/do previously learnt languages play in the lexical production of the subjects' written German?

This study attempts to describe the *Interlanguage*³ (IL) of Kenyan learners of German as a foreign language, and is of particular fascination, not only because of the nature of learner language in general, but because of the multilingual background, the previously learned and acquired languages of the subjects involved. In this particular paper we discuss the issue of lexical interlanguage transfer as one form of the cross-linguistic influence manifested in the German IL of the subjects under investigation.

A multilingual learner corpus in the sense of this study has been taken to have a twofold meaning: (1) the subjects under investigation are individuals who speak at least three languages, and are currently learning a fourth i.e. German as a foreign language, and (2) the corpus consists of texts from learners. We thus speak of the multilingual learner and the learner corpus. The corpus contains an IL in which the

¹ L4 in this study is used merely to describe and throw light of the acquisition order of the languages of the subjects involved. Hammarberg (2001) for example, would describe the language under study as L3.

² Institut für Deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin
e-mail: victoria.anyango@googlemail.com

³ The term *Interlanguage* was introduced by Selinker (1972) to refer to learners' versions of the target language.

history of all the previously learned languages of the subjects in question is contained, and which is a language variety in its own right.

It has been assumed that bi- or multilingualism is at least as frequent in the population of the world as pure monolingualism, perhaps even more frequent. The number of people who speak only their mother tongue or, apart from the mother tongue, speak just one other foreign language, are relatively few in comparison to those who have learnt more than one foreign language (Krumm, 2001). Mackey (1967) discusses why individual bilingualism is bound to be a common situation in the world. He points to the multitude of small linguistic communities, the wide currency and usefulness of national and international languages and people's increasing mobility across language borders.

The type of research carried out from the early 1940s to the 1980s focused however heavily on second language acquisition involving the constellations of two languages: the mother tongue L1 and one foreign language L2. The assumption that second, third, fourth foreign languages were not worthy of further investigation arose at a time when results of studies into language acquisition processes were assumed to be true of whatever language that had been acquired or learnt, regardless of whether the language in question was the second, third or fourth language. Recent research however, has shown that although second language acquisition has a lot in common with multilingual acquisition, there are some differences regarding complexity and diversity. The learning of L3, L4, L5, L_x (x ≥ 2) is characterised for example by several transferability possibilities, occurring because of the presence of several languages (Hufeisen 2003).

Assuming that humans are potentially polyglot⁴ by nature, an adequate theory of language competence, use and acquisition should be able to account for polyglot cases, and preferably take these as the norm, treating pure bi- or monolingualism as special cases. The theory will have to take into account that the (linguistically mature) individual may normally have two or more languages to handle, and that: (a) the speaker is able to choose according to intention which language to use, (b) the speaker's languages can regularly be kept apart, but also get mixed or influence each other, and furthermore (c) that the person's competence in the various languages will normally not be at equal levels. In view of this, the language acquisition process in polyglots offers itself as an interesting field of study, particularly the ways in which the individual's languages interact in such complex cases."

(Hammarberg 2001: 22)

In Section 2 the sociolinguistic situation of the multilingual subject in Kenya shall be described and the different language systems that make up the multilingual individual, or the trilingual, shall be explained. Section 3 shall deal with the role of previously learned languages in the production and acquisition of another, and in the last section the learner corpus, as well as sample findings, shall be described.

⁴ Hammarberg (2001) defines a polyglot as a person with knowledge of three or more languages.

2. The multilingual learner

2.1 The sociolinguistic situation of Kenya

Kenya is a typical representation of a multilingual society. It is however difficult to state the exact number of languages spoken in Kenya depending upon the source one is citing and whether or not one is referring to only grammatically stable codes⁵ (Ogechi 2003). A report of the Constitution of Kenya Review Commission (CKRC) (2000: 95) puts the number of languages at 70 while other sources (e.g. Webb and Kembo-Sure 2000) put it at 42. Ogechi (2003) reports that if the number of grammatically unstable but widely used codes among the youth – Sheng – is included, then the count of languages spoken in Kenya goes up.

The languages of Africa are generally divided into four major phyla: Niger-Congo, Nilo-Saharan, Afroasiatic, and Khoisan (Childs 2003: 21). Three of these groups are represented in Kenya: the Bantu languages belong to the Niger-Congo family, the Nilotic languages to the Nilo-Saharan group and the Cushitic to the Afroasiatic (Schladt 1997).

English is also regarded as one of the Kenyan languages (Webb and Kembo-Sure 2000: 13), and is the official language used in government, international business, diplomacy and in the school education system. Kiswahili is the national language and is also used for government administration as well as casual inter-ethnic communication.

2.2 Types of Multilingualism

When two languages are involved in the acquisition process we only have two possible acquisition orders: the second language can be acquired either after the L1 (L1→L2), or at the same time as the L1 (Lx + Ly).

In the case of third language acquisition there is greater diversity and there are at least four possible acquisition orders (Cenoz 2000). The three languages can be acquired consecutively (L1→L2→L3). Other possibilities include the simultaneous acquisition of two languages (Lx/Ly) that could take place after the L1 has been acquired (L1→Lx/Ly) or before the L3 is acquired (Lx/Ly→L3). Another possibility involves simultaneous contact with three languages (Lx/Ly/Lz). The diversity of the possible acquisition orders can increase when the acquisition of four languages is considered, and would reflect the different situations of second and multilingual acquisition taking into account the simultaneous or consecutive acquisition of the different languages. This diversity is increased if we consider that the acquisition process can be interrupted by the process of acquiring an additional language and then restarted again (L1→L2→L3→L2). The subjects in this study could be said to have the following language acquisition order: L1→Lx/Ly→L4. L1 would be the mother tongue; Lx and Ly represent Kiswahili and English, L4 German. German as a foreign language is offered as an optional subject in some secondary schools, the majority of the schools offer French as a foreign language. The lessons comprise of three lessons

⁵ Ogechi (2003) has defined a grammatically stable code as one that has native speakers and one whose grammar and lexicon are fairly stable and can be studied while a grammatically unstable code is one whose lexicon and grammar are unstable. Examples of the latter include Sheng, which is widely spoken among the urban and a few rural youth in Kenya and sounds like Kiswahili but has a distinct and unstable vocabulary.

of forty-five minutes per week. If one decides to pursue German at the university level, one has a choice of a degree in German Studies or an educational degree in German as a foreign language.

The Kenyan learners are brought up, generally speaking, in a trilingual community, and the degree to which they need to use the possible languages in the community depends on whether they live in a town or upcountry environment. In the towns, members from all the different language communities are usually represented, especially in the capital and major cities. One would thus on average speak Kiswahili, English and/or Sheng within the wider community⁶, and the ethnic language at home within the family environment, if at all. It is not uncommon to find that even at home, Kiswahili and English or Sheng are the languages of communication. Within an upcountry environment, the language spoken in the home is usually the same as the language of the community dominant in that area, and in this regard the use of Kiswahili or English is restricted to business communication or to use within an educational context. I suggest the following broad classification of the Kenyan trilingual subject, based on Hoffmanns (2001) classification of the trilingual subject, which takes into account both the circumstances and the social context under which the subjects become users of three languages:

- (1) A trilingual child in a major town environment is brought up with one home language, which is different from the community languages represented in that town, and/or which is not dominantly spoken. He/she acquires the second languages before going to school;
- (2) A child in an upcountry environment is brought up with one home language, which is the language dominantly spoken in the wider community, and acquires the second languages in the school context.

Clearly, as Hoffmann (2001) remarks, there are many trilinguals who straddle (such) categories or move in and out from one to another⁷. It is also possible to find a situation in which a child is brought up with *two* home languages. The Kenyan trilingual subject is thus usually a member of a heterogenous language environment, and learns to speak at least three or four languages in order to meet his/her different communicative needs.

Hoffman (2001) goes on further to describe trilingual language competence as containing the linguistic component – knowledge of the three language systems – and also the pragmatic component, which consists of sociolinguistic, discourse and strategic competences pertaining to the three languages involved. In addition, it includes the ability to function in bilingual or trilingual contexts, which requires decisions on code choice and code-switching. In other words, trilingual competence enables trilinguals to create their own linguistic means in order to master particular communicative situations. It is unusual though, for trilinguals to have a stable pattern

⁶ Kikuyu, a Bantu language, is more often spoken by both native and non native speakers in the capital city; one of the reasons for this could be the close geographical distance to Central province, where the kikuyu community lives.

⁷ Mombasa for example, is classified as one of the major cities in Kenya, but because of the large presence of native speakers of Kiswahili, this would be the language dominantly spoken within the community, perhaps within the business and school context as well. Here one also has the interesting phenomenon of “beach boys”, who have (because of varied business interests) contact with tourists, and who thus roughly speak their languages. It is quite usual to find them speaking some German or Italian, without any formal schooling in the language.

of similar competencies in their three languages, but this does not generally affect their overall linguistic competence.

An issue related to language competence concerns the notion of speech modes in trilinguals. Grosjean (1992) proposes that we should see the bilingual's speech modes in terms of end points on a scale, and I shall adapt his metaphor to describe the trilingual subject in the study. The Kenyan trilingual moves from his/her monolingual speech mode when talking to a person of the same ethnic community, and so leaves the other languages deactivated, then changes along the continuum to a bilingual / trilingual speech mode where she/he makes use of both / all three of her/his languages in the form of frequent switching and borrowing. He/she thus may use any one of seven different possible language constellations

The consideration of the interaction of the languages used by the subjects under study leads us to ask about the role that these play in their production and acquisition of a foreign language, in this case German.

3. The roles of L1 and L2 in L3 Production and Acquisition

3.1 Third Language or L4?

The notion of third language (L3) is used in Tertiary Language Acquisition research in a sense that relates to the established notions of first and second language (L1, L2). Languages that are acquired after the first language (or first languages, in the case of infant bilingualism) are commonly termed second languages; a person may acquire one or more L2s. Hammarberg (2001) uses the term L3 for the language that is currently being acquired, and L2 for any other language that the person has acquired after L1. L3 in this technical sense is not necessarily equal to language number three in order of acquisition.

For the purposes of description we have used L4 to refer to German, which, determined in chronological terms, is acquired by the subjects under study after their native language, Kiswahili and English. We shall refer to English or Kiswahili as L2, when speaking of IL transfer from either of these to the German IL.

There have been several studies that provide evidence that that prior L2s have a greater role to play than has usually been assumed (see e.g. Chandrasekhar, 1978, Hufeisen, 1991, Hammarberg and Williams, 1993). Various factors that condition L2's influence on L3 have been proposed. Thus many studies provide evidence for a factor of *typological similarity*: influence from L2 is favoured if L2 is typologically close to L3, especially if L1 is more distant. Other factors include *proficiency*: L2 influence is favoured if the learner has a high level of competence in the L2, and if the L2 has been acquired and used in natural situations, and *recency*: an L2 is activated more easily if the speaker has used it recently and thus maintained easy access to it. A further factor is that of *L2 status*: there appears to be a general tendency to activate an earlier secondary language in L3 performance rather than L1 (Hammarberg 2001).

3.2 Interlanguage and Interlanguage Transfer⁸

Selinkers (1972) *Interlanguage Hypothesis* assumes that ILs are natural languages, systematic throughout their development. ILs reflect learners' attempts at constructing a linguistic system that progressively approaches the target system. By definition, ILs are incomplete, intermediate and in a state of flux, and are the products of interaction between two linguistic systems, namely those of L1 and L2, or, as in the case of a multilingual individual, L_x ($x \geq 2$).

The influence of a non-native language on another non-native language - i.e. the documented transfer from one interlanguage to another – is referred to as *interlanguage transfer* (de Angelis and Selinker 2001). They assume that all linguistic systems present in the speaker's mind may be simultaneously interacting and competing in IL production. Previous language transfer theories were highly restricted, being primarily based on the interaction between two language systems, the native and one non-native system. For IL transfer to occur however, more than two linguistic systems must be present in the speaker's mind.

Technically, interlanguage transfer cannot occur without a minimum of three linguistic systems. Viewing interlanguage transfer as involving at least three linguistic systems would allow us to capture the simultaneous interaction, and importantly the possible competition, between more than two linguistic systems at a single point in time.

(de Angelis and Selinker 2001: 44)

IL transfer can occur at several levels, and its extent varies. In no other area, however, is the importance of psychotypological factors, perceived similarities, more in the foreground than lexis (Ringbom 2001). For this study we shall differentiate different types of lexical transfer as defined by Ringbom. Lexical transfer can be manifested as transfer of form or transfer of meaning, and in the present study I shall present a sample of the transfer found in the German IL of the subjects under investigation.

4. The Learner Corpus

The aim of this study, to investigate the German IL of Kenyan learners of German as a foreign language, has partly been accomplished by the creation of a learner corpus consisting of essays from nineteen adult subjects. Learner corpora are defined by Granger (2002) as follows:

Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogenous way and documented as to their origin and purpose.

The learner corpus examines Tertiary Language Acquisition, where Tertiary is taken to mean L_x ($x \geq 2$). It is hoped that close its investigation can highlight a range

⁸ Herdina and Jessner (2002) use the term cross-linguistic interaction to include not only transfer and interference but also code switching and borrowing phenomena and it thus serves as an umbrella term for all the existing transfer phenomena.

of features of IL in order to understand the system underlying it and concurrently or subsequently compare the IL with one or more native speaker norms in order to assess the extent of the deviation.

The area of linguistic enquiry known as learner corpus research, which has only existed since the late 1980s, has created an important link between the two previously disparate fields of corpus linguistics and foreign/second language research. Using the main principles, tools and methods from corpus linguistics, it aims to provide improved descriptions of learner language which can be used for a wide range of purposes in foreign/second language acquisition research and also to improve foreign language teaching... the very nature of the evidence it uses makes it particularly powerful methodology, one which has the potential to change perspectives on language...

(Granger 2002)

My learner corpus is a part of the essay corpus Falko (Fehlerannotiertes Lernerkorpus) that contains texts of advanced learners of German as a foreign language⁹. It is being developed at the Humboldt-Universität zu Berlin, and the texts come from students with varied native speaker backgrounds.

The learners of German in this study have previous knowledge of at least three other languages (their mother tongue, Kiswahili and English), and speak these with varying degrees of fluency, as described above. With the use of a questionnaire, information was gathered about their linguistic skills and learning history.

The subjects were asked to do a C-Test in order to determine their proficiency in German, and only texts from subjects who had attained a particular level¹⁰ of proficiency in German were used. They were asked to write a timed composition on any one of four topics, and were not allowed to use any reference material. The argumentative essay topics were taken from ICLE (International Corpus of Learner English)¹¹ with their kind permission.

4.1 Sample Results and Discussion

The following examples illustrate the two types of lexical IL transfer under discussion found in the subjects' German IL production, and are based on Ringboms (2001) descriptions of the different types of transfer of form (as affecting linguistic forms) and transfer of meaning (as affecting the semantic pattern). For the purposes of this study we have restricted ourselves to two types of transfer, and present these in the following table adapted from Ringbom:

⁹ Falko is freely available at the following website: <http://www2.hu-berlin.de/korpling/projekte/falko/>

¹⁰ Learners who scored at least 60% and more on the C-Test. This is one of the requirements that learner texts in Falko should meet.

¹¹ ICLE was developed at the Louvain Centre for English Corpus Linguistics. www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/icle.htm

Type of Transfer		Underlying Cause	Transfer of form or meaning	From which language	Examples
A	Language switch and coinage	Insufficient awareness of intended linguistic form, instead of which (a modified form of) an L2 word is used	Form (results in non-existing TL word)	L2 English	... diesselben <i>Ministers</i> finden sich in große <i>Scandals</i> ...
B	Totally or partial deceptive cognate	Awareness of an existing L3 form, but confusion caused by formal similarity to a word in another language	Form and Meaning	L2 English	...aus Länder, die <i>Zivilkrieg</i> erfahren... (Engl. civil war)

Table 1: Examples of lexical interlanguage transfer

In the Type A transfer, the learner has tried to produce a German word, but a formally similar English word has been produced instead. This is a transfer of form, not meaning. In the transfer error of type B, the learner assumes that what is a homonym or polysemous word in the German IL has a meaning corresponding to what is the core meaning of the equivalent L2 word. There is also confusion caused by the formal similarity to the word *Zivil* in German. In this case a transfer of meaning *and* form seems to have taken place.

5. Conclusion

The question of why and under what circumstances transfer from the native language system, or from any other system, is blocked (or permitted) in favour of a particular IL when other options may be available to the speaker seems here to be answered by the concept of the psychotypological relation between L2 and L3.

Languages perceived to be similar (roughly = related) to the target language naturally provide many more reference points for the learner than do wholly unrelated languages.

(Ringbom 2001: 65)

The subjects' native languages and Kiswahili, as non Indo-European languages, seem at this level of the investigation to play less of a role in the lexical IL transfer than English. It could also be that the subjects' higher level of competence in English and the fact that they use it very often in their daily affairs¹² played a role in favouring the transfer of lexical items from English rather than from their mother tongues or Kiswahili.

As mentioned above, this is a sample of the preliminary results of the investigation of German IL among the multilingual subjects involved. Further investigation of the corpus is taking place, and it is hoped that further aspects of the cross-linguistic interaction between the language systems in the multilingual learners as manifested in their German IL will come to light.

¹² It is to be stressed that the level of competence and frequency of use of the different languages are obtained at this stage of the investigation from the subjects' own self evaluation.

References

- Cenoz, J. (2000) Research on multilingual acquisition, in J. Cenoz and U. Jessner (eds) *English in Europe: The Acquisition of a Third Language*. Clevedon: Multilingual Matters.
- Chandrasekhar, A. (1978) 'Base Language'. *International Review of Applied Linguistics* 16, 62–65.
- Childs, T. G. (2003) *An Introduction to African Languages*. Amsterdam: John Benjamins Publishing Co.
- Constitution of Kenya Review Commission 2002. *The Main Report of the Constitution of Kenya Review Commission*. 18 September, 2002.
- De Angelis, G. and Selinker, L. (2001) Interlanguage Transfer and Competing Systems in the Multilingual Mind, in J. Cenoz, B. Hufeisen and Jessner Ulrike (eds) *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters.
- Granger, S. (2002). A Bird's-eye view of learner corpus research, in S. Granger, J. Hung and S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 3–33. Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Grosjean, F. (1992) Another view of bilingualism, in R. Harris (ed.) *Cognitive Processing in Bilinguals*, pp. 51–62. Amsterdam: Elsevier Science Publishers.
- Hoffmann, C. (2001) The Status of Trilingualism in Bilingual Studies, in J. Cenoz, B. Hufeisen and U. Jessner (eds) *Looking beyond Second Language Acquisition Studies in Tri- and Multilingualism*, pp. 13–28. Tübingen: Stauffenberg.
- Hammarberg, B. and Williams, S. (1993) A study of third language acquisition, in B. Hammarberg (ed.) *Problem, Process, Product in Language Learning* pp. 60–70.
- Hammarberg, B. (2001) Roles of L1 and L2 in L3 Production and Acquisition, in J. Cenoz, B. Hufeisen and U. Jessner (eds) *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters.
- Herdina, P. and Jessner, U. (2002) *A Dynamic Model of Multilingualism. Perspectives of Change in Psycholinguistics*. Clevedon: Multilingual Matters.
- Hufeisen, B. (1991) *Englisch als erste und Deutsch als zweite Fremdsprache. Empirische Untersuchungen zur fremdsprachlichen Interaktion*. Frankfurt: Peter Lang.
- Hufeisen, B. (2003) Modelle der wissenschaftlichen Beschreibung von individueller Mehrsprachigkeit und Bedingungen ihrer Entstehung, in R. de Cillia, H-J. Krumm and R. Wodak (eds) *The Cost of Multilingualism. Globalisation and Linguistic Diversity*, pp. 57–70. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Krumm, H-J. and Jenkins, E-M (eds) (2001) *Kinder und ihre Sprachen – lebendige Mehrsprachigkeit. Sprachenporträts*. Wien: Eviva
- Mackey, W. F. (1967) *Bilingualism as a World Problem*. Montreal: Harvest House.
- Ogechi N. O. (2003) 'On Language Rights in Kenya'. *Nordic Journal of African Studies. Volume 2 Number 3*. Helsinki: Helsinki University Printing House.
- Ringbom, H. (2001) Lexical Transfer in L3 Production, in J. Cenoz, B. Hufeisen and U. Jessner (eds) *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters.

- Schladt, M. (1997) Kognitive Strukturen von Körperteilvokabularen in kenianischen Sprachen. Köln: Afrikanistische Monographien 8.
- Selinker, L. (1972) 'Interlanguage' *International Review of Applied Linguistics* 10, 209–231.
- Webb, V. and Kembo-Sure (2000) African Voices; An Introduction to the Language and Linguistics of Africa. Cape Town: Oxford University Press Southern Africa.