Typological and social constraints on language contact: Amerindian languages in contact with Spanish
Gomez Rendon, J.A.

Citation for published version (APA):
Chapter 4

The Research Program

This chapter outlines the research program underlying the present investigation and discusses a number of questions related to the methods used in the collection and analysis of data. The chapter is organized in three sections. The first section reviews studies on linguistic borrowing and the methodologies used for the investigation of this linguistic phenomenon. I outline a program of research on linguistic borrowing oriented to solving methodological shortcomings on the basis of an in-depth analysis of three contact situations in different areas of Latin America which involve one donor language (Spanish) and three typologically different languages (Paraguayan Guaraní, Ecuadorian Quichua and Mexican Otomí). The choice of languages is substantiated on a theoretical and methodological basis. The third section sets the main questions guiding this research and how I intend to provide answers to them. Afterwards I discuss the general hypotheses to be tested on the corpora collected for the aforementioned languages. The hypotheses are developed from the premise that the typology of the languages in contact co-determines the degree and the form of lexical and grammatical borrowing and their functions in the recipient language (cf. section 3.3). The hypotheses will be further developed for each of the investigated languages in the following chapters. The last section addresses a number of methodological issues concerning the process of data collection, the setup of the corpora and their characteristics, the representation of data and the criteria used in the statistical analysis. The discussion of some research problems and their solutions rounds off the chapter.

4.1 A critical overview of studies on linguistic borrowing

Linguistic borrowing was first studied in the nineteenth century as part of comparative and historical linguistics. Early studies viewed linguistic borrowing as a random phenomenon influenced by countless non-linguistic motivations which, albeit interesting, lacked relevance for linguistic theory. A notable exception is the classical study on “language mixture” by Whitney (1881) in which he sketched a number of regularities of linguistic borrowing. In the early twentieth century Meillet (1921), Vočadlo (1938) and others advanced the research on linguistic borrowing in more precise terms. However, the most important step toward a systematic study of borrowing was taken by Haugen (1950). Haugen not only addressed current issues in the field but provided a thorough classification of borrowings which somehow survives up to date. The second breakthrough in the study of linguistic borrowing is the work of Thomason and Kaufmann (1988). According to Myers-Scotton (2002:
the innovative aspects of their proposal are: a) the distinction between interference and borrowing; b) the linking of borrowing types to intensity of contact through a borrowing scale; and c) the inclusion of lexical and structural borrowing in one scale. Their contribution was certainly decisive in establishing linguistic borrowing as a central issue in contact linguistics.

Most of the aforementioned studies do not grow out of corpus-based investigations. Instead, they are collections of findings from different case studies. Many of these studies are not comparable on account of the differences in social, cultural and historical aspects of the contact settings, but also because the methods followed in the collection and analysis of data are not standardized. Case studies on linguistic borrowing concern individual pairs of donor and recipient languages in different contact situations all over the world. They are based on corpora gathered from oral or written sources. A serious shortcoming of case studies is their preference for analyzing linguistic borrowings from dominant (usually European) languages in native languages. This bias is explained by the fact that lexical and grammatical borrowing in colonial and neocolonial settings has a greater impact on native languages, but also by the focus on imperfect learning and acquisition in the study of non-standard varieties of dominant languages, such as Spanish, Portuguese or English in the Americas. A further shortcoming of case studies is that sociolinguistic information concerning bilingualism and other speaker factors are not considered in the analysis. Moreover, an important number of case studies make no functional distinction between code-switching and borrowing. As I have shown, the distinction between both phenomena is important in diglossic situations because it implies different levels of bilingualism and distinct communicative answers to the pressure exerted by contact. Finally, the great majority of case studies on linguistic borrowing have not been conducted in a theoretical framework that provides the researcher with tools for analysis and hypotheses about expected outcomes in specific situations.

The findings of an increasing number of case studies are used to support distinct views on language contact (cf. Romaine 1989; Thomason 2001; Myers-Scotton 2002) and different scales of borrowability (cf. Haugen 1953; Muysken 1981; Singh 1981; Bakker and Hekking 1999; Field 2003), their results being often interpreted in an ad hoc manner. It has been repeatedly claimed, for example, that borrowings are grouped in recognizable word classes and certain word classes are preferred over others. Findings that confirm these claims are numerous in the literature, but much less numerous are studies which explain these claims in the framework of linguistic theory. The fact that authors resort to ad hoc explanations has resulted in a larger number of borrowing scales without sufficient theoretical foundations. The

---

1 A notable exception for the treatment of data and the consideration of several intervening factors is Stolz and Stolz (1996).
Theories on Linguistic Borrowing

Universals of language contact proposed by Moravcsik (1978) are an exception, not only because of the typological principles underlying her proposal but also because of their arrangement in a coherent set. Still, it was not difficult to demonstrate that these universals are disconfirmed by evidence from many contact situations around the world. Other explanations are less consistent though potentially promising: e.g. frequency and transparency as factors influencing the preferential borrowing of open classes over closed classes. In addition, the study of the influence of typology on contact-induced language change has a long tradition in linguistics (Meillet 1921; Vočadlo 1938; Weinreich 1968). All in all, explanations of linguistic borrowing are not theory-driven and therefore lack a systematic treatment.

4.1.1 Studies on linguistic borrowing in Latin America

Studies on linguistic borrowing in Latin America have focused on the influence of Spanish (and Portuguese) on Amerindian languages. Some of these studies are based on isolated examples extracted from grammars and dictionaries, while others result from a corpus-based investigation. In both cases findings are not discussed in a coherent framework.

One of the earliest reports on linguistic borrowing in Amerindian languages is Boas (1931), who presents an inventory of Spanish elements in Modern Nahuatl (Uto-Aztecan). More recent studies on Nahuatl are Hill and Hill (1986) and Flores Farfán (1999), both of which show the impact of Spanish-origin borrowings on the lexicon and the grammar of the native language up to the emergence of mixed sociolects. Studies on lexical borrowing for other languages of the Uto-Aztecan family from the perspective of lexical acculturation are Silver and Miller (1997) for Mountain Pima and Comanche, and Campbell (1987) for Spanish influence on Pipil.

From an ethnolinguistic point of view Brown (1994) provides an interesting survey of Spanish and English lexical acculturation in Native American languages.

For the influence of Spanish on Otomí (Otomanguean) there are several reports by Hekking and Bakker (1998, 1999, and 2007) plus a comprehensive study by Hekking (1995) on language shift and restructuring in the Otomí dialect of Santiago Mexquititlán. These studies are corpus-based and typology-oriented and stress the relevance of typological factors for borrowing. Further studies on native Mexican languages are Brody (1976) on Spanish-origin particles borrowed as discourse

---

2 The influence of Amerindian languages on the local and regional varieties of Spanish is comparatively less studied. Worthy of mention are Haboud (1998) on Ecuadorian Andean Spanish influenced by Quichua; Knvoshein de Canese and Corvalán (1987) and Dietrich (1995) on Paraguayan Spanish influenced by Guaraní; and Flores Farfán (1998; 2000) on Mexican Spanish influenced by Nahuatl and learned as a second language by native speakers of native Mexican languages. Still, none of these studies are corpus-based and only few are theory-driven. The result is that most of their findings remain at a purely descriptive level.
markers in various Mayan languages; and Knowles-Berry (1987) on linguistic decay in Chontal (Mayan) as a result of Spanish borrowing. Both authors do not base their findings on corpora but on isolated data collected through elicitation techniques.


The contact between Spanish and Guaraní in Paraguay has remained relatively understudied despite its relevance for the relation of bilingualism and borrowing. Until recently all the studies of the influence of Spanish on Guaraní were due to one single author. Morínigo (1936) is an extensive compilation of Spanish words in Guaraní accompanied with examples, glosses and ethnographical explanations. Morínigo (1959) and Morínigo (1982) address the phenomenon of Spanish grammatical borrowing in Guaraní from two slightly different perspectives. A recent corpus-based study of lexical and grammatical borrowing in Paraguayan Guaraní from a typological perspective is Gómez-Rendón (2007b). While many sociolinguistic studies insist on the unique condition of Paraguay as the only bilingual country in Latin America, systematic analyses of the impact of Spanish on Guaraní are scarce and limited to isolated cases.

### 4.2 A program of research on linguistic borrowing

Considering the theoretical and methodological shortcomings of most studies in linguistic borrowing it was therefore necessary to set up a research program that solves these deficiencies by

a) identifying a set of constants and variables in order to have control over the data and arrive at valid language-specific and cross-linguistic conclusions;
b) working from a corpus-based perspective in order to base the analysis on realistic data; and
c) interpreting the data in the framework of a linguistic theory that predicts the behavior of borrowing in different typological and sociocultural settings.
The research on linguistic borrowing which is reported in this book follows these guidelines. In the following I explain how the guidelines translate into a set of parameters.

**Borrowing versus shift-induced interference**

This investigation focuses exclusively on borrowing and the ways Spanish influences Amerindian languages. The study of the influence of these languages on Spanish through imperfect learning is not considered here. The focus on borrowing implies that all the speakers interviewed are native speakers of one Amerindian language while their second language, either learned or acquired, is a variety of Spanish spoken in their area of origin or residence. Accordingly, the native language is the dominant language of the speech community to which the speakers belong, while Spanish is used mostly in transactions with the mainstream society, usually though not always, outside the borders of the speech community. The Spanish proficiency of the speakers interviewed varied depending on such variables as gender, age and formal schooling.

**Borrowing versus code-switching**

Spanish-origin elements present in the native language are either borrowings or code-switches. Each type has its own linguistic features and may be distinguished from the other according to several parameters discussed in section 2.5.2.1. Since my purpose is to identify constraints on the borrowing of Spanish elements in the native language and their accommodation to the recipient language, only the first type of elements (borrowings) is considered for analysis. Code-switches were identified and labeled according to their length (single-word or complex) but they were not included in the analysis. In order to assess the overall contributions of code-switching and borrowing in the corpus of each language, texts were analyzed for number, length and type of code-switches. This helped us measure the differential contributions of foreign elements in the samples and identify relations between them. Likewise, it helped us establish a distinction between single-word borrowings and single-word code switches, on the one hand, and between complex borrowings and complex code switches, on the other.

Considering the controversial nature of single-word code switches and their relation to borrowing (section 2.5.2.2) it was necessary in several cases to conduct a fine-grained analysis of the phonological and morphosyntactic integration of the foreign elements in the native discourse before assigning them to either category.

---

3 Arguably, nonce borrowings represent an intermediate stage between established borrowings and code switches, thereby forming a continuum between codeswitching and borrowing.
This procedure also helped us make a clearer distinction between complex borrowings (e.g. frozen constructions lexicalized as single units in the recipient language) and complex code switches (e.g. chunks of foreign material inserted in the recipient language). I do not pretend in this way to settle the issue of the distinction of both types of mixing phenomena. My purpose was purely instrumental to the extent that the application of several criteria of structural accommodation could help us draw a dividing line between borrowings and code switches for analysis.

*Lexical and grammatical borrowing*

Different from a number of case studies on linguistic borrowing that focus on the lexicon, this study includes both lexical and grammatical elements. This inclusion seeks to integrate the findings in one single model of linguistic borrowing. This model aims at explaining linguistic borrowing as the outcome of contact between two typologically different languages on the basis of four sets of constraints determined by a) the principle of functional explanation; b) the principles of system compatibility and incompatibility; c) the scales of borrowability; and d) the parts-of-speech systems of the languages in contact. Each constraint produces individual hypotheses that predict the higher or lower probability for a linguistic element to be borrowed from one language provided certain conditions are met. The treatment of lexical and grammatical borrowing within a single model is based on the premise that each type of borrowing is determined by time and intensity of contact and other sociocultural motivations. It implies a continuum stretching from lexical to grammatical borrowing as time and intensity of contact increase. While the view of borrowing as a continuum is based on the concept of ‘scale of borrowability’, the matching of borrowing types with degrees of contact is inspired by the borrowing scale proposed first by Thomason and Kaufmann (1988) and refined later by Thomason (2001).

*Typologically different languages as recipients of borrowing*

Most studies on linguistic borrowing are investigations of individual cases in different parts of the world. Only a few analyze borrowing in cross-linguistic perspective. This perspective is required when the purpose is to find structural constraints on the borrowing process as in the present investigation. In order to find cross-linguistic constraints on borrowing it is therefore necessary to analyze more than one language but also typologically different languages. This procedure helps us deal with the immense variety of languages of the world. While it is realistically unfeasible to collect and analyze all the languages representing the world’s typological variety, it is clearly viable to begin with a sample of languages whose typological profiles are different from each other.
The typological criteria for the selection of the languages of this study include language family, morphology, lexical classes, types of affixation, types of adpositions, and word order. The languages selected were Guaraní, Quichua and Otomí. These languages meet the conditions of typological variation along the aforementioned parameters. Each language belongs to a different family, though all of them are spoken in the Americas. Guaraní is a Tupi language of the Tupi-Guaraní family, spoken by five million people in Paraguay and the Argentinean province of Corrientes. Quichua is a language of the Quechua family, spoken by one million people in the Andean Highlands of Ecuador. Otomí belongs to the Otomanguean branch of the Otopamean language family and is spoken by three hundred thousand people in different states of Central Mexico. As regards their morphological typology, Guaraní is originally a polysynthetic language while Quichua is typically agglutinative and Otomí more analytic than the other two at the level of the sentence but synthetic at the level of the phrase (Bakker et al 2008). Also, the three languages differ from each other in their systems of parts of speech: Guaraní and Quichua are flexible languages, but the former shows a larger number of word classes which are used predicatively; Otomí, on the other hand, is a rigid language without adjectives. The type of affixation in these languages makes them different too: Guaraní has both prefixes and suffixes (Gregores and Suárez 1967); Quichua has only prefixes (Cole 1982); and Otomí uses both plus numerous clitics (Hekking 1995). Guaraní and Quichua, on the contrary, are typically postpositional languages; Otomí diverges from them in that it uses prepositional constructions to link elements within the noun phrase. Basic word order is another point of divergence among these languages. Guaraní shows a relatively free word order, with a preference for SVO order (Gregores y Suárez 1967). Otomí has VOS and VS as basic orders, although the frequency of SVO constructions is increasing in usage today (Lastra de Suárez 1994; Hekking 1995). Finally, Quichua is a typical verb-final language, even though the occurrence of SVO constructions has increased over the last years as a result of contact with Spanish. A detailed characterization of the typology of these languages is presented in Chapters 6 through 8.

In addition to linguistic reasons, the investigation of these languages offers good fieldwork conditions because a) they are vital in their number of speakers; b) they are spoken in vast areas of their respective countries; and c) they are relatively well described, with a number of grammars and dictionaries, some of which date from the first years of the Spanish colonization.

*One language as the source of borrowing*

A fundamental methodological premise of the present investigation is that the recipient languages must be typologically different in order to produce cross-linguistically valid conclusions, and the source language must be kept constant for
all the recipients and typologically different from them. These conditions are met satisfactorily by Spanish.

As a result of the worldwide expansion of the Spanish Empire in the fifteenth and sixteenth centuries, the Spanish language is present in the five continents today. Except for the territory of today’s Brazil, which became part of the Portuguese Empire in the early fifteenth century, Spanish was the official language in Central and South America during the three centuries of Spanish colonization. In addition, Spanish was also spoken in several areas in the southern United States until the end of the nineteenth century. The linguistic heritage of Latin American republics reflects the dominance of the Spanish language, and the countries in which the investigation was conducted are no exception. Spanish remains the official language and the largest in terms of speakers in Ecuador, Paraguay and Mexico. Spanish has a long history of contact with hundreds of native languages all over Latin America. At the same time, Spanish has remained typologically distinct from native languages in spite of their substratum and adstratum influence. Spanish remains a fusional-synthetic language with prepositions and flexible word order, and distinct lexical classes for individual syntactic functions. In all, the sociolinguistic and linguistic conditions of Spanish in the Americas allow for the investigation of linguistic borrowing from one language into typologically different languages. In this way the input to borrowing is kept constant and the foundations are laid for cross-linguistic generalizations.

Spanish America as a sociocultural region

In my model of contact-induced language change the ultimate motivations for borrowing are essentially nonlinguistic. Therefore, any cross-linguistic analysis of linguistic borrowing requires that sociocultural motivations be similar enough to allow for comparison. Keeping the sociocultural variable constant is unfeasible in any realistic study given the enormous variety of national, regional and local societies and cultures found in a vast region like the Americas. Nevertheless, by taking Spanish America as the geographical space for the contact between the source language (Spanish) and the native languages, an important degree of social and cultural unity is warranted. This unity is substantiated by a series of historical events and the resulting sociolinguistic facts.

4 Arguably, the Spanish varieties spoken in Ecuador, Mexico and Paraguay are not the same. In fact, it is possible to find a number of lexical and morphosyntactic differences. However, these dialects remain mutually intelligible and typologically similar to each other and to Peninsular Spanish. While this is not the case of Spanish varieties spoken by non-native speakers – like many forms of Indian Spanish described in the literature (e.g. Flores-Farfán 2000) – these were not considered for the present investigation.
The Spanish rule in the Americas lasted over three centuries and left sociocultural imprints in the continent. The identification of Spanish America as a cultural region is based on the cultural heritage shared by all Spanish-speaking countries in the continent. This heritage expresses in a number of facts, from legal and administrative apparatuses to architecture and religion. Notice that a focus on similarities does not neglect differences, which are many and very important. Differences are firmly rooted in the heritage of numberless Indian cultures, many of which survive to the present and became the basis for the foundation of nations such as the Guaraní in Paraguay, the Inca in Peru or the Aztec in Mexico.

As regards the language, the great majority of countries in Spanish America are diglossic societies. In this context Spanish is the language of prestige while the native languages are usually excluded from public spheres. This is the case of Quichua, Otomí and even Guaraní in their respective countries. The official status of Guaraní in Paraguay does not make it different from other native languages in sociolinguistic terms.

In sum, the sociocultural heritage of all Spanish-speaking countries in Latin America and the condition of dominance of Hispanic culture over native cultures allow a controlled comparison between the contact situations analyzed in this book.

A corpus of spontaneous speech from a representative group

Not being based on the investigation of a corpus, most studies of linguistic borrowing take as material for their analysis a collection of isolated examples from the languages in question or a sample of written texts extracted from other sources. The approach of this study is the opposite. I have analyzed individual corpora for the three languages of the sample. These corpora were collected in situ according to a number of criteria to be explained in section 4.4.1. No elicitation was used in the process and speech events were recorded in socially and culturally relevant settings. In doing so I sought to reduce speech monitoring and de-contextualization of verbal exchanges to the minimum. Speech monitoring is an important factor influencing the number and type of borrowings in situations where the source language is used by speakers of the recipient language for their interaction with people from outside of their speech communities, especially if the languages are in a diglossic situation. Accordingly, the corpus of each language is comprised of spontaneous speech in face-to-face interactions.5

A further criterion for the setup of the corpora was the inclusion of a representative group of speakers from the speech communities under study. By including lectal variation into the sample I could chart the speech of different

---

5 In few cases, however, second-hand material from other sources was to be used in order to cover other registers or sociolects to which the researcher could not have access.
subgroups: men and women, older and younger generations, and literate and illiterate speakers. The rationale was twofold: the requirement of representativity of the sample; the integration of the time variable in contact-induced language change. The latter criterion requires some comment here. Because changes in language are in most cases gradual and take place within the time span of several generations, it is necessary for any study of language contact to plot changes diachronically as well. Ideally, only a longitudinal study over a time span of several decades would meet this requirement. Because such a study is out of the question here, the next option was to simulate time in the sample through charting the speech of individuals from different age groups. This procedure enabled us to find, for example, that older generations, usually more conservative than younger ones in their linguistic usage, prefer borrowing over code-switching. The findings were supported by historical information from earlier sources, when available, which confirmed the gradual entrance of foreign elements in the form of code switches as the intensity of contact and bilingualism increased.

The collection and analysis of corpora meeting the criteria of sociolinguistic and diachronic representativity are time-consuming tasks. In our case, the samples required between up to fifty nine speakers depending on the language and the average text length normally surpassed 1500 tokens. Accordingly, the resulting corpora differed in size, from 60,000 to 110,000 tokens.

**Contact-induced language change in diachronic perspective**

Linguistic borrowing as a contact-induced language change is a process and calls for a diachronic view. The process is visible in the way foreign elements are incorporated into the recipient language: from their occurrence in the idiolects of bilingual innovators to their subsequent spreading among other speakers and finally to the speech community as a whole. The process is also reflected in the gradual accommodation of foreign elements when used over a longer period of time: from their non-assimilation at the phonological level to their full integration into the phonological system of the recipient language. In the absence of similar corpora for previous stages of the language, the process can be mapped, to a certain extent, either by recording the speech of individual speakers from different age groups, as argued above, or by studying earlier sources in the form of grammars and dictionaries. For the languages of the sample there exist no pre-contact corpora that serve as a yardstick for comparison. Instead, we have a series of linguistic descriptions prepared since the early years of the Spanish conquest by members of the clergy for evangelization purposes. The availability of grammatical descriptions is not the same for the three languages, however. Fray Pedro Cáceres and Alonso Urbano wrote the first grammatical description and the first dictionary of Otomí in 1580 in 1605, respectively. Fray Ruiz de Montoya published a grammar and
dictionary of Guaraní only in 1640. The sources appeared even later for Ecuadorian Quichua: the first grammatical sketch was published only in the mid-eighteenth century (Anonymous 1760) while the first dictionary came up a few years later (Velasco 1787). All these sources were used as a reference for earlier stages of the language and served to keep track of early borrowings in the languages.

A theoretical framework providing analytic tools and testable hypotheses

As indicated at the beginning of the chapter, most studies of linguistic borrowing do not base their analysis of data on a specific theoretical framework. Their interpretation of data is obscured by the use and abuse of ad hoc linguistic explanations without previous hypotheses about the expected number, type and use of borrowings. The present study seeks to fill this gap by working within the theoretical framework of linguistic typology and sociolinguistics and avail of their respective analytic tools and hypotheses. The concepts and implications of the theoretical framework were amply discussed in Chapter 3. In the following section I develop several hypotheses by taking as a point of departure the premise that the typology of the languages in contact co-determines the degree and the form of lexical and grammatical borrowing.

4.3 Research questions and general hypotheses

On the basis of the model of contact-induced language change developed in section 2.6 I assume that nonlinguistic motivations are the primary cause of linguistic borrowing in any contact situation and the outcomes of contact thus motivated are modeled by linguistic and nonlinguistic factors and conditions. Theoretically, any contact-induced change is possible provided that a number of nonlinguistic circumstances are met. Still, research on language contact shows that not all possible changes are attested and that the outcomes of contact are regular and systematic to a great extent. Therefore, the central question to be answered is how regularities in contact-induced language change are influenced by structural factors derived from the typological features of the languages in contact. For this purpose I investigate the number, type and functional adaptation of Spanish lexical and grammatical borrowings were investigated in three typologically different Amerindian languages (Guaraní, Quichua and Otomí). The research questions may be detailed as follows:

- Do linguistic factors play a role in the borrowing process of Spanish elements into Guaraní, Quichua and Otomí, and if so, to what extent and under what conditions?
- More specifically, do the typological profiles of these languages play a role in the borrowing process, and if so, to what extent and under what conditions?
And even more specifically, do the lexical and grammatical categories of these languages play a role in the borrowing process, and if so, to what extent and under what conditions?

In other words, the main goal is to identify the linguistic factors that promote or inhibit borrowing of certain lexical and grammatical categories and the linguistic conditions that speed up or slow down borrowing. The influence of linguistic factors and conditions will be confronted with the influence of nonlinguistic motivations in each contact setting. Generally speaking, the interaction between nonlinguistic motivations and linguistic factors and conditions is expressed in the following terms:

Native speakers of language R (recipient) who also speak language S (source) with different degrees of proficiency are driven by nonlinguistic circumstances to incorporate a lexical or grammatical feature of their second language (S) into their native language (R). This feature either is available in language R or not. In the first case borrowed features either replace an already existing feature in language R or make it more specific. In the second case, the paradigm of features in R is either extended or adapted by the entrance of features from S. The chance for any feature from S to be borrowed by speakers of R is co-determined by nonlinguistic and linguistic factors. Linguistic factors are the typological similarity between R and S, the equivalence between word classes in R and S, the frequency of borrowed features in S, the translatability of features from S, the paradigmaticity of word classes in R, etcetera. Similarly, chances for any feature from S to be borrowed more rapidly by R are increased by the frequency of the native feature being replaced with a feature from S.

The influence of nonlinguistic motivations and linguistic factors and conditions on the outcomes of contact is expressed in the form of hierarchies. For linguistic borrowing these hierarchies represent arrangements of lexical and grammatical elements ordered according to the higher or lower probability of borrowing from one language to another. An extensive discussion of hierarchies or scales of borrowability was presented in Chapter 3. In the following I present a number of hypothesis based on: a) the Principle of Functional Explanation; b) the Principle of System Compatibility; c) the scales of borrowability; and d) the Hierarchy of Parts of Speech. Each hypothesis is based on a hierarchy of linguistic factors or elements and will be tested on the corpora of the languages.

4.3.1 Borrowing hypotheses from the Principle of Functional Explanation

From the Principle of Functional Explanation (cf. 2.6.2) the principal hierarchy of factors governing the borrowing process is H.1: pragmatic factors are the most
decisive in any contact situation, followed by semantic and formal ones. In this perspective, pragmatic and semantic factors are promoters of borrowing while formal factors act mainly as constraints.

**H.1**

<table>
<thead>
<tr>
<th>Pragmatic factors</th>
<th>Semantic factors</th>
<th>Syntactic-Morphological-Phonological factors</th>
</tr>
</thead>
</table>

This hierarchy of factors is translated in more specific terms by positioning pragmatic markers on the top of the scale in relation to other linguistic elements. This is expressed in subhierarchy 1.1 below. Discourse markers include basically topic and focus markers.

**H.1.1**

<table>
<thead>
<tr>
<th>Discourse markers</th>
<th>Other linguistic elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic marker</td>
<td>Focus marker</td>
</tr>
</tbody>
</table>

**4.3.2 Borrowing hypotheses from the Principle of System Compatibility**

A second hierarchy predicts the probability that a foreign element may be borrowed easier than others. The hierarchy is based on the classification of morphemes in free and bound. The prediction states that free forms are more prone to borrowing than bound forms in a contact situation. The term ‘morpheme’ is a cover term including not only grammatical forms such as inflectional or derivational affixes but also free lexical morphemes such as nouns or adjectives. Free morphemes and bound morphemes are roughly equivalent to lexical and grammatical classes, respectively. Exceptions are function words, which are free grammatical morphemes. Some clitics also share characteristics of free and bound morphemes and therefore occupy an intermediate position in the hierarchy.

**H.2**

<table>
<thead>
<tr>
<th>Free morpheme</th>
<th>Clitic</th>
<th>Bound morpheme</th>
</tr>
</thead>
</table>

Because languages have different morphological profiles, this hierarchy is insufficient to account for all cases of linguistic borrowing. It is therefore necessary to include the morphological type of the languages in contact on the basis of the Principle of System Compatibility. According to this principle, “any form or form-meaning set is borrowable from a donor language if it conforms to the morphological possibilities of the recipient language with regard to morphological structure” (Field 2002: 42). In other words, if the recipient language is fusional-
synthetic, it may borrow virtually any foreign element, including free forms such as independent words and bound forms such as roots, agglutinating and fusional affixes. If the language is isolating-analytic, it may borrow only free forms while most bound forms (roots and affixes) are not borrowable in principle. In these terms subhierarchy H.2.1 below makes the predictions of H.2 more specific by establishing which languages have fewer difficulties in borrowing elements from another language on the basis of their morphological type:

**H.2.1**

| Fusional-synthetic | > | Agglutinating-synthetic | > | Isolating-analytic |

### 4.3.3 Borrowing Hypotheses from the Scales of Borrowability

The split between lexicon and grammar is the basis of a third hierarchy. This is at the same level of H.1, and both are considered to interact with each other. The hierarchy orders lexical and grammatical elements according to their degree of borrowability. From the borrowing scales discussed in Chapter 3, lexical elements are represented as more borrowable than grammatical elements and occupy the first place:

**H.3**

| Lexical elements | > | Grammatical elements |

An extension of this hierarchy concerns class type. Classes are grouped in open, half-open and closed depending on the number of elements and the possibility that other (foreign) elements are incorporated. The incorporation of borrowings to closed classes is more difficult than their incorporation to half-open and open classes. In general terms open classes and closed classes correspond to lexical and grammatical elements, respectively. Half-open classes are a matter of content because they are halfway between lexicon and grammar. The open-closed constraint on classes is known as paradigmaticity (cf. 2.6.2.2). This is summarized in the following subhierarchy:

**H.3.1**

| Open class | > | Half-open class | > | Closed class |

A further subhypothesis extends the predictive value of H.3 by specifying the degree of borrowability of lexical and grammatical classes. Subhypothesis H.3.2 derives from the scales of borrowability discussed in section 3.5. To visualize the relation
between the central hypothesis and its subhypotheses they have been conflated into one single hierarchy with subhierarchies in descending order:

**H.3.2**

<table>
<thead>
<tr>
<th>Lexical elements</th>
<th>&gt;</th>
<th>Grammatical elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open class</td>
<td>&gt;</td>
<td>Half-open class</td>
</tr>
<tr>
<td><strong>Noun &gt; Verb &gt; Adjective &gt; Adverb &gt; ...Adpositions... &gt; ...Auxiliary &gt;...&gt; Article</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The basic prediction from H.3.2 is that elements of open lexical classes are more borrowable than elements of closed grammatical classes. The subhierarchy has two interpretations depending on the language: a) from the perspective of the source language it implies that elements of this language which correspond to major parts of speech are in general easier to borrow than elements which correspond to other parts of speech; b) from the perspective of the recipient language it implies that the parts of speech of this language determine the borrowability of a foreign element – which may or may not correspond to an equivalent class in the source language – depending on the openness or closedness of the target class. Both interpretations derive from the hypothesis that the systems of parts of speech of the languages involved in the borrowing process are relevant to determine the type of borrowings. I consider the second interpretation more relevant to the analysis pursued here because it determines not only the possibility for a foreign element to enter in a certain class but also how this element is used in the recipient language. It is clear that the typological profile of both languages in contact is relevant, but that of the recipient language is decisive.

**4.3.4 Borrowing hypotheses from the Parts-of-Speech Theory**

The theory of parts of speech developed by Hengeveld (1992) and Hengeveld et al (2004) defines parts of speech primarily on syntactic grounds and considers the phrase as the basic syntactic unit. Phrases can be referential (noun phrase) or predicational (verb phrase). Each phrase is composed of two slots, one for heads and one for modifiers. Along these parameters the theory establishes the existence of three types of languages: flexible languages, with one lexical class for two or more syntactic functions; differentiated languages, with one lexical class for one syntactic function; and rigid languages, with no lexical classes for one or more syntactic slots and morphosyntactic strategies used instead.

This classification of languages according to parts of speech has an important consequence for lexical borrowing: the types of lexical items that may be borrowed in a given contact situation depend, among other things, on the flexibility or rigidity of the languages in contact. Accordingly, the general hypothesis is formulated as follows: