Mannheim papers in multilingualism, acquisition and change
https://doi.org/10.25521/mapmac.2021.159

Student Edition 2021
Vol. 3, pp. 1–72

Interlanguage as a projection of abstract lexical structure from the native and the target language: Testing the Composite Matrix Language model of second-language acquisition

Anne Apel

Mannheim University

apel.anne94@gmail.com

Abstract

This thesis investigates the Composite Matrix Language model of second-language acquisition (Jake, 1998) which predicts which structures may arise in the construction of interlanguage independently of the specific language pair involved. The model views second-language acquisition as an instance of language contact where the native and the target language interact at an abstract level of lexical structure preceding surface-level projections. The abstract lexical structure of lexical items is stored in the mental lexicon and includes information on three levels, namely 1) lexical-conceptual structure, 2) predicate-argument structure, and 3) morphological-realization patterns.

The Composite Matrix Language model argues that in interlanguage construction, abstract lexical structure from the native and the target language is split up and recombined, resulting in a composite morphosyntactic frame projecting surface structure. Based on its psycholinguistic role in interlanguage construction, the model constrains the possible contribution of the native language. This thesis tests whether the predictions of the model can account for interlanguage structures in English learners of French, i.e., whether these can be explained as projections of composite lexical structure from the native and the target language.

For this purpose, spoken interlanguage data from ten British sixth-form students in their sixth year of learning French (*Newcastle Corpus*, Myles & Mitchell,

2013b, available via the *French Learner Language Oral Corpora* database, Myles & Mitchell, 2013a) are analyzed.

The results of the analysis largely confirm the predictions stated under the Composite Matrix Language model. Ambiguous and problematic structures are discussed with regard to their implications for the predictions of the model. In general, the thesis provides evidence for the universal principles of language contact assumed under the Composite Matrix Language model.

Keywords: French L2-acquisition; morphosyntactic transfer; L1 influence; Matrix Language Frame model; learner corpus analysis

Contents

1 Introduction **3**

2 Theoretical background **6**
2.1 Second-language acquisition 6
2.2 Transfer—Definition and historical overview 6
2.3 Interlanguage 8

3 The Composite Matrix Language model of second-language acquisition **8**
3.1 The Matrix Language Frame model 9
3.1.1 Assumptions and principles under the Matrix Language Frame model 10
3.1.2 The 4-M model 12
3.1.3 The Abstract Level model 15
3.1.4 Criticism of the Matrix Language Frame model 17
3.2 The Composite Matrix Language Model 18
3.2.1 Theoretical foundations and goals of the model 18
3.2.2 Assumptions and predictions of the Composite Matrix Language model 18
3.2.3 Summary: The Composite Matrix Language model 22
3.3 Literature review 23
3.3.1 Application of the Matrix Language Frame model to language-contact phenomena 23
3.3.2 Evidence for the Composite Matrix Language model 27
3.4 Research questions 31

4 Method **33**
4.1 Data 33
4.1.1 Participants 33
4.1.2 Tasks 34
4.1.3 The subcorpus analyzed in this study 35
4.2 Analysis 35

5 Results and discussion 37
5.1 The target language as the basis of interlanguage surface structures 37
5.2 Interlanguage surface structures as projections of composite lexical structure from the native and the target language 40
5.2.1 Native language abstract lexical structure filling gaps in interlanguage 40
5.2.2 Insertion of elements from the target language into interlanguage gaps 42
5.2.3 Open gaps in interlanguage 50
5.3 Ambiguous cases and potential counter-examples 51
5.3.1 Unclear status of morpheme type under the 4-M model 52
5.3.2 Ambiguous structures 54
5.3.3 Structures not explicable through either the native or the target language 55
5.3.4 Native language structure affecting late system morphemes in interlanguage 56

6 Conclusion and outlook 60

References 64

Appendix 70

List of abbreviations 72

1 Introduction

Describing, predicting, and explaining the developmental paths learners take when they are acquiring a second language (L2) has been an intriguing challenge for several generations of researchers in the field of second-language acquisition (SLA). More than four decades ago, Selinker (1972) coined the term of interlanguage (IL) referring to the learner variety developing in the acquisition process. Based on the observation that the vast majority of second language learners, although attempting to produce sentences conform to a particular L2 norm, do not reach native-like competence in their target language (TL), he assumed IL to be a separate linguistic system developing in SLA that is influenced by both the learner's native language (L1) and the TL (Selinker, 1972, p. 214; see also chapter 2.3).

Until today, most SLA research has focused on the acquisition of grammar, i.e., syntax and morphology, which is seen as central to the process of learning an L2 (Mitchell, Myles, & Marsden, 2013, p. 6). In this regard, much research has been conducted to investigate the role of transfer from the L1 in IL development. However, the descriptive focus of research, mainly concentrating on the observation of IL surface structures and developmental paths but failing to account for these, has been criticized repeatedly (cf., e.g., Jake, 1998; Wei, 2015). One model that claims explanatory power for the development of IL structures is the 'Composite Matrix Language (CML) model' of second-language acquisition by Jake (1998) linking SLA to language contact and transfer at an abstract level preceding surface structures.

The CML model aims to constrain which structures may arise in IL development and what the participating languages may contribute. It views the acquisition of a second language as an instance of language contact where the L1 and TL interact in the construction of interlanguage. Therefore, Jake (1998) argues that the process of SLA should be explicable through principles governing other language-contact phenomena as well, positioning her model within the framework of the 'Matrix Language Frame (MLF) model' (Myers-Scotton, 1993, 1997, 2002; see also chapter 3.1).

The MLF model was originally developed to explain and predict the structural outcomes of codeswitching and later extended to account for grammatical structures in language contact in general, claiming that all language-contact phenomena are governed by the same abstract principles (Myers-Scotton, 2002, p. 6). It is based on the psycholinguistic finding that language production is driven by abstract lexical entries in the mental lexicon which mediate between the conceptualization of a message and its grammatical and phonological encoding by projecting surface structures (Levelt, 1989). The central hypothesis stated under the MLF model is that the differential activation of the languages participating in language-contact situations determines what they may contribute to bilingual speech (Myers-Scotton, 1993, 2002).

The CML model of SLA assumes that learners intend to construct IL from abstract lexical structure from TL as much as possible (Jake, 1998). As described in the Abstract Level model related to the MLF model, abstract lexical structure refers to the complex information stored in the lexical items in the mental lexicon and contains three levels: 1) lexical-conceptual structure, i.e., semantic information, 2) predicate-argument structure concerning the assignment of thematic roles and how they are mapped onto syntactic structure, and 3) morphological-realization patterns involving surface-level devices such as word order and agreement morphology to mark hierarchical relations (Myers-Scotton & Jake, 1995; see also section 3.1.3). Due to the incomplete acquisition of the TL, the abstract lexical structure of TL lexical items is not always fully available to the learners. In this case, abstract lexical structure underlying lexical items in the L1 may fill in gaps in TL lexical items, representing transfer at an abstract level. That is, the CML model predicts that in the construction of IL, abstract lexical structure from L1 and TL is split up and recombined, resulting in a composite structure projecting actual IL surface structures (Jake, 1998).

However, the L1 is constrained in what it may contribute as predicted by the 4-M model of morpheme classification which is related to the MLF model as well (Myers-Scotton & Jake, 2000; see also section 3.1.2). The 4-M model categorizes different types of morpheme based on how they are accessed in language production and has implications for the differential distribution of these morpheme types in mono- and bilingual speech. It differentiates morphemes that are activated at a conceptual level on the basis of their semantic information in order to convey speaker intentions, such as verbs or nouns, from morphemes such as case and agreement markers that are structurally assigned at a later level when syntactic constituents are assembled and hierarchical relations are spelled out. According to Jake (1998), only abstract lexical structure underlying conceptually activated L1 morphemes may be transferred to IL, and it may only be inserted into gaps in IL that are projected by conceptually activated morphemes. Furthermore, it is claimed that the predictions made under the CML model generalize

to SLA data in general independently of which languages are involved.

Up until today, evidence for the predictions stated under the CML model has remained illustrative and mainly comes from a corpus of interlanguage data from Chinese and Japanese learners of English (Wei, 1995) which is not representative (see section 3.3.2). This thesis tests whether the predictions also apply to other L1-TL pairs. Specifically, it investigates whether the CML model of SLA by Jake (1998) can explain IL surface structures observed in learners with the L1 English and the L2 French. This superordinate question is divided into four more specific questions targeted on the concrete predictions of the model:

1. Are the IL surface structures observed in the learner data based on TL lexical items?
2. Can the IL surface structures observed in the data be explained as projections of composite lexical structure from the L1 and TL?
3. Do the restrictions for the limited influence of the L1 hold?
4. Are there IL structures that cannot be accounted for within the framework of the CML model, and if yes, what are their implications for the predictions of the model?

Crucially, this thesis aims to provide a comprehensive examination of the learner data, which does not only provide further illustrative evidence but also addresses IL structures that may seem problematic for the predictions of the CML model.

The thesis is structured as follows. Chapter 2 provides definitions of the concepts of second-language acquisition, transfer, and interlanguage. Chapter 3 introduces the theoretical underpinnings of the thesis. Chapter 3.1 gives an overview of the principles and hypotheses stated under the MLF model as the theoretical foundation for the CML model (section 3.1.1). It also introduces the 4-M model of morpheme classification and the Abstract Level model (sections 3.1.2 and 3.1.3) adding precision to the predictions of the MLF model. Chapter 3.1 ends with several comments on criticism of the MLF model (section 3.1.4). In chapter 3.2, the CML model of SLA is outlined. After introducing the foundations and goals of the model (section 3.2.1), the specific assumptions and predictions for the construction of interlanguage are described (section 3.2.2). At the end of chapter 3.2, a brief summary of the CML model is given (section 3.2.3). Chapter 3.3 reviews the current state of research concerning the relevance of the principles of the MLF model for language-contact phenomena in general (section 3.3.1). It also gives an account of previous research aiming to provide evidence for the predictions made under the CML model and identifies the limitations of these studies with regard to their representativeness (section 3.3.2). On the basis of the discussion of the literature, chapter 3.4 develops the specific research questions guiding the analysis in this thesis. Chapter 4 describes the data and the method of analysis. In chapter 5, the results of the analysis are presented following the order of the research questions. Chapter 5.1 and 5.2 demonstrate that the predictions of the CML model can be confirmed on the basis of the learner data, illustrating that TL is the basis of IL surface structures and that IL structures can be explained as projections of composite lexical structure from L1 and TL. Following that, chapter 5.3 addresses ambiguous data and discusses potential counter-examples. Finally, a summary of the results is given in chapter

6 which also includes comments on the implications of the findings and open questions for further research.

2 Theoretical background

2.1 Second-language acquisition

In general, the concept of 'second-language acquisition' refers to "the process of learning another language after the native language has been learned" (Gass, 2013, p. 4). This process involves both the creation and testing of hypotheses based on the input available to the learner and the use of knowledge of a native language or other previously learned languages (Gass & Selinker, 1993, p. 6).

In a broad sense, the term 'second language' may refer to any non-native language that is learned after the L1 has been learned, even if it is a third or fourth language (Gass, 2013, p. 4; Mitchell et al., 2013, p. 1). However, the analysis in this thesis is restricted to situations of SLA where the L2 is the first and only non-native language of the participants, that is, the learner data analyzed here come from learners who do not know any further L2s than French. This is because the existence of more than one previously learned language makes it difficult to determine with certainty whether the source of non-target-like IL structures is the L1 or a previously learned L2 (see chapter 4).

Depending on the framework, the term 'acquisition' may also be used in a broad or a more narrow sense. In some definitions, it includes all types of learning situations, referring to both formal, systematic classroom-based learning and informal learning through natural exposure (Gass, 2013, p. 4; Mitchell et al., 2013, p. 2). Some theories of SLA, however, differentiate between SLA and 'foreign language learning'. In this case, SLA refers to situations where a non-native language is learned in an environment where the TL might act as a language of wider communication, i.e., where learners are exposed to high amounts of TL in their everyday lives (Gass, 2013, p. 5; Mitchell et al., 2013, p. 1). Foreign language learning, on the other hand, denotes situations where the L2 is learned as a truly foreign language in the environment of the native language, usually in classroom settings with significantly lower amounts of exposure to natural TL (Gass, 2013, p. 5; Mitchell et al., 2013, p. 1).

The data analyzed in this thesis come from learners who have acquired the L2 as a foreign language in a classroom setting and, thus, qualify as instances of foreign language learning. However, according to the claim of universal applicability of the CML model of SLA, its predictions should hold independently of whether an L2 has been acquired in a classroom or a more naturalistic setting. Therefore, both kinds of learning are subsumed under the general term SLA in this thesis.

2.2 Transfer—Definition and historical overview

In SLA research, the notions of 'transfer' or 'cross-linguistic influence' synonymously refer to the study of the role of the L1 in the development of IL (Gass, 2013, p. 79; Odlin, 2003,

p. 436). Specifically, transfer is viewed as "the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (Odlin, 1989, p. 27; as cited in Odlin, 2003, p. 436). While the definition by Odlin includes influence from previously learned L2s, this thesis only focuses on the influence of the L1 in SLA.

Cross-linguistic influence occurs on all linguistic levels, e.g., phonology, morphology, syntax, semantics, pragmatics, and orthography (Odlin, 2003, p. 437). Crucially, transfer from the L1 does not necessarily result in non-target-like structures in IL. Therefore, positive and negative transfer are differentiated as two possible outcomes of the same process. Positive transfer occurs when reliance on the L1 leads to target-like IL utterances in cases where L1 and L2 are similar. In contrast, reliance on L1 results in non-target-like IL structures when L1 and L2 are different, representing negative transfer (Gass & Selinker, 1993, p. 6).

Also, a distinction has to be made between transfer on the surface level versus on abstract levels, related to different hypotheses on what kind of L1 knowledge influences L2 development. Thus, transfer may either involve surface forms or abstract structures and processes (Foley & Flynn, 2018, p. 97). In the CML model to be tested in this thesis, it is assumed that transfer takes place at an abstract level of lexical structure underlying actual surface morphemes (Jake, 1998; see also chapter 3.2).

The role of the L1 has been an important research area in the investigation of SLA from early on. In the 1950s and 1960s, 'Contrastive Analysis' (E.g., Fries, 1945; Lado, 1957) was the predominant paradigm, focusing on the structural difference between L1 and L2 and the consequences on the outcome of SLA (Thomas, 2018, p. 30). The proponents assumed a strong influence of the L1 on second-language acquisition, resulting in correct L2 utterances where L1 and L2 are similar and in errors when L1 and L2 are different (Thomas, 2018, p. 30; VanPatten & Williams, 2015, p. 20).

From the 1970s onwards, counterevidence to these predictions accumulated, indicating that not all of the predicted errors occur and that not all of the errors found in learners' productions can be attributed to the L1 (Mitchell et al., 2013, pp. 35–36; VanPatten & Williams, 2015, p. 22). Additionally, studies investigating the order of morpheme acquisition in learners with different L1 backgrounds (Bailey, Madden, & Krashen, 1974; Dulay & Burt, 1973, 1974, 1975; Dulay, Burt, & Krashen, 1982; Meisel, Clahsen, & Pienemann, 1981) gave hints at the existence of a potentially universal order, implying that the L1 plays a rather minor role as one of many factors in the process of SLA (VanPatten & Williams, 2015, p. 22).

These findings led to an increased interest in the language actually produced by learners (Mitchell et al., 2013, p. 35). One of the first researchers to focus on the study of learners' errors is Corder (1967) drawing attention to the nature of the linguistic system emerging in SLA (Mitchell et al., 2013, p. 36) which led to the emergence of the term interlanguage by Selinker (1972).

From the 1980s onwards, different sub-disciplines have emerged in SLA research. For example, formal accounts of SLA related to generative grammar are primarily concerned with the structure of the mental system underlying syntactic knowledge in individuals, focusing on the questions whether L2 learners have access to Universal Grammar and whether parameters

are transferred or re-set in SLA (e.g., Schwartz & Sprouse, 1996; Vainikka & Young-Scholten, 1996; White, 1985). In contrast, cognitive, functionalist, and sociolinguistic approaches deal with the transition mechanisms in the process of SLA, considering factors such as frequency and the readiness of learners (Ellis, 2003; Krashen, 1981, 1985; Pienemann, 1984, 1998). Generally, the existence of cross-linguistic influence on all linguistic levels is widely accepted. Transfer represents a central concept in every theory or model of SLA as well as in related fields of linguistic research, such as language contact and language universals (Odlin, 2003, p. 437; Odlin & Yu, 2016, p. 8). However, the extent and nature of influence from the L1 is still under debate (Mitchell et al., 2013, p. 16; Odlin & Yu, 2016, p. 8; Thomas, 2018, p. 28).

2.3 Interlanguage

The notion of 'interlanguage' emerged as a result of the increased interest in the linguistic system developing in SLA after the rejection of Contrastive Analysis. In a seminal article, Selinker (1972) observes that while the vast majority of L2 learners do not attain native-like competence in their L2, their productions are based on specific and systematic rules (which are not identical to those of the TL). He suggests that the language originating in the process of SLA is a complete linguistic system in its own right that evolves over time and coins the term 'interlanguage' to refer to this system (Selinker, 1972, p. 214).

Also, Selinker notes that IL contains deviations from the TL that remain stable regardless of the amount of instruction or exposure to TL. He refers to this phenomenon as 'fossilization' (Selinker, 1972, p. 215). These instances of non-native-like attainment may be related to motivational aspects: In SLA, learners seek to be able to communicate meaning. Thus, the acquisition of grammatical elements may stop when learners feel that this major goal is achieved (Selinker, 1972, p. 217).

The terms and definitions introduced in this chapter are essential for a discussion of any model of SLA. The next chapter presents the CML model of SLA which relates transfer to language contact and approaches L1 influence at an abstract level.

3 The Composite Matrix Language model of second-language acquisition

The CML model is based on the assumption that SLA is a language-contact phenomenon structured by the same principles that also govern other language-contact situations (Jake, 1998, p. 361). This chapter intends to give an overview of the theoretical background of the model, its concrete assumptions and predictions, and the state of the art, leading to the formulation of the research questions that guide this thesis.

In chapter 3.1, the MLF model is introduced. This model represents the theoretical underpinnings of the CML model of SLA. Also, the 4-M model of morpheme classification and the Abstract Level model complementing the MLF model are presented because their implications are relevant for the CML model as well. Furthermore, criticism of the MLF model is considered.

In chapter 3.2, the specific assumptions of the CML model derived from the MLF model and its sub-models are outlined. Chapter 3.3 reviews research into the applicability of the MLF model to language contact phenomena and discusses previous work providing evidence for the predictions made under the CML model. Following that, chapter 3.4 develops the specific research questions for the analysis.

3.1 The Matrix Language Frame model

The MLF model (Myers-Scotton, 1993, 1997, 2002) was first developed to explain and predict the structural outcomes of codeswitching as a form of bilingual speech. In this case, 'bilingual speech' refers to any type of contact phenomena involving utterances that contain surface morphemes and/or abstract lexical structure from more than one language (Myers-Scotton, 2002, p. 7). 'Codeswitching', in turn, is defined as a specific type of bilingual speech where elements from one language are inserted into utterances from another language during the same conversation (Myers-Scotton, 1993, p. 3).

The MLF model differs from other structural approaches to codeswitching and language contact. Previous models have often remained descriptive, limiting their constraints to linear surface relations (e.g., Poplack, 1980) and working with traditional lexical categories (e.g., Timm, 1975). In contrast, the MLF model seeks to not only label different structures or constraints but also to explain them. Hence, the principles and constraints stipulated in the MLF model are motivated by psycholinguistic insights into mono- and bilingual language production coming from research into bilingual speech, speech errors, and aphasia (Myers-Scotton, 1993, p. 46, 2002, pp. 13–14).

A central hypothesis of the MLF model is based on Levelt's (1989) model of the speech production process. Levelt states that language production is driven by abstract lexical entries in the mental lexicon which mediate between the conceptualization of a message and its grammatical and phonological encoding. These lexical entries are conceived of as complex entities stored in the mental lexicon and containing information on several levels: semantic and syntactic information subsumed under the term 'lemma information', or short, 'lemma', as well as morphological and phonological information, also called 'form information' (Levelt, 1989). The MLF model adopts the view that these abstract lexical entries in the mental lexicon underlie actual surface morphemes and that their lemma information, when activated on the basis of their meaning, points to their syntactic information for the construction of surface structures (Myers-Scotton, 1993, p. 49, 2002, pp. 14–15). This process is assumed to be divided into several, presumably incremental stages. Firstly, at the conceptual level, speaker intentions are mapped onto language by the activation of semantic/pragmatic feature bundles that best match the message the speaker wants to express. Secondly, at the lemma level, these semantic/pragmatic feature bundles select lexical entries on the basis of their semantic meaning from the mental lexicon. At the level of the Formulator, the syntactic information contained in these lemmas results in morphosyntactic building procedures generating surface structures (Myers-Scotton, 2002, pp. 23–25).

The MLF model is especially concerned with the structural outcomes of language contact

in individuals, i.e., in the minds of bilingual speakers (Myers-Scotton, 2002, p. 4). It takes account of both the overt and covert effects of bilingualism. 'Overt' effects are observable at the level of surface structure, e.g. in codeswitching (Myers-Scotton, 2002, p. 1). 'Covert' effects, in contrast, are instances of bilingual speech where all surface morphemes come from one language but contact takes place at a more abstract level, e.g., grammatical influence in language attrition or transfer of abstract lexical structure in L2 acquisition (Myers-Scotton, 2002, p. 3).

Furthermore, while acknowledging that sociolinguistic factors may have an influence on the actual occurrence of codeswitching and contact phenomena in general, the model mainly focuses on contact at the level of morphosyntax (Myers-Scotton, 2002, p. 4). Accordingly, the unit of analysis under the MLF model is the bilingual complementizer phrase (CP)[1] (Myers-Scotton, 2002, p. 54) containing constituents from more than one language or mixed constituents containing morphemes from more than one language (Myers-Scotton, 2002, p. 56). In this sense, the bilingual CP may be defined as a large mixed constituent in itself (Myers-Scotton, 2002, p. 58).

The MLF model in its original version only applies to one particular type of codeswitching labeled 'classic codeswitching'. In this type of bilingual speech, only one of the participating languages provides the abstract grammatical structure that underlies surface utterances (Myers-Scotton, 2002, p. 8). In its later versions, the model has been extended to other language-contact situations and used to explain grammatical structures in contact phenomena in general (Myers-Scotton, 2002). In the model, 'contact phenomena' are defined as the different structural outcomes of bilingual speech in the languages involved, such as creole formation, bilingual language acquisition, attrition, language shift, and the development of interlanguage in second-language acquisition (Myers-Scotton, 2002, p. 7).

The central argument underlying the extension of the MLF model from classic codeswitching to other language-contact phenomena is the claim that there exists a small number of abstract principles which govern all contact phenomena, "even though the details of how they are played out in the various phenomena may differ" (Myers-Scotton, 2002, p. 6). That is, even in cases where the abstract grammatical structure underlying surface utterances does not come from only one of the participating languages, the same constraints that are at play in classic codeswitching apply. This type of bilingual speech is labeled 'composite codeswitching' where not only the surface morphemes come from more than one language but the abstract grammatical structure underlying surface utterances is a result of convergence of abstract structure from both participating languages (Myers-Scotton, 2002, p. 8).

3.1.1 Assumptions and principles under the Matrix Language Frame model

The first premise of the MLF model is the assumption that the morphosyntax of any CP is structured by an analyzable or resolvable frame (Myers-Scotton, 2002, p. 8). In bilingual speech,

[1] The CP is the highest level of projection in a tree of syntactic structures projected by a complementizer. Thus, the complementizer represents the head of the clause and has the IP (inflection phrase) as its complement containing smaller constituents such as determiner phrases (DPs), verb phrases (VPs), and prepositional phrases (PPs) (Radford, 2009, p. 50).

the abstract grammatical frame is called the 'matrix language' (ML). This frame is an abstract configuration prior to surface structures, including specifications of morpheme order and directions for the realization of functional elements (Myers-Scotton, 1993, p. 76). It is important to note that the term 'matrix language' is introduced in the model to refer to the abstract morphosyntactic frame of a bilingual CP. Crucially, the ML is not to be confused with one of the participating languages. Rather, it is viewed as an abstraction of one of the participating languages that serve as its source (Myers-Scotton, 2002, p. 66). Still, the term is also used in the MLF model to refer to the source language itself for ease of reference (Myers-Scotton, 2002, p. 67).

The premise of the ML structuring bilingual CPs is closely related to the 'uniform-structure principle' and the 'asymmetry principle for bilingual frames'. The uniform-structure principle proposes that constituents in any language have a uniform structure and that the structural requirements for any given constituent must be observed whenever it appears (Myers-Scotton, 2002, p. 8). The asymmetry principle for bilingual frames takes up on the idea that the languages of a bilingual speaker are not activated equally (Grosjean, 1988). As a consequence of this "morphosyntactic dominance of one variety in the frame" (Myers-Scotton, 2002, p. 9), it is the structural requirements of the dominant language that always have to be met in bilingual speech (Myers-Scotton, 2002, p. 8).

Taken together, these premises result in the first out of two crucial hierarchies established under the MLF model, namely the 'matrix language-embedded language opposition'. It refers to the abstract level of linguistic competence and represents the idea that the bilingual's languages do not contribute equally in bilingual speech due to their unequal levels of activation (Myers-Scotton, 2002, pp. 15–16). The ML is the language that is more salient psycholinguistically with regard to the production of mixed constituents and thus contributes more. In this case, 'more' does not necessarily refer to the number of morphemes, even though this is often the case. Rather, it indicates that the ML provides "*more abstract structure and structure of a certain type*" (Myers-Scotton, 2002, p. 15, author's emphasis). Crucially, as the more dominant language, only the ML sets the grammatical frame for bilingual speech (Myers-Scotton, 1993, p. 6, 2002, p. 16), i.e., the grammatical procedures underlying the production of mixed constituents are based on the ML exclusively (Myers-Scotton, 1993, p. 83). The other, less dominant language largely supplies content morphemes that are integrated ('embedded') into the ML frame and is therefore called the 'embedded language' (EL) (Myers-Scotton, 2002, p. 16).

Finally, another asymmetry marking bilingual speech is summarized under the 'morpheme-sorting principle'. It states that from a psycholinguistic perspective, there exist different types of morpheme, which, due to their differences on an abstract level of language production, "do not have equal possibilities of occurrence" in bilingual speech (Myers-Scotton, 2002, p. 9).

The morpheme-sorting principle leads to the stipulation of the second central hierarchy under the MLF model, namely the 'content morpheme-system morpheme opposition', distinguishing two different types of morpheme. The opposition operates at an abstract level as well. It recognizes that surface morphemes and structures are based on abstract lexical items, taking into account how these are organized in the mental lexicon and that they are accessed at different stages in the language production process (Myers-Scotton, 2002, p. 16). This, in

turn, affects how they participate in bilingual speech (Myers-Scotton, 2002, p. 16). In this opposition, 'content morphemes' are lexical elements that convey the semantic and pragmatic meaning of a message. 'System morphemes', in contrast, are related to constituent structure and indicate relationships between content morphemes, largely (but not exactly) fitting the class of functional elements (Myers-Scotton, 2002, p. 15).

According to Myers-Scotton (1993, p. 6), the content morpheme-system morpheme opposition is used by the ML as the major organizing device in setting the morphosyntactic frame of bilingual utterances. In this scenario, the status of EL morphemes determines whether they may appear in mixed constituents or not. This assumption is stated in the 'system-morpheme principle' defining a critical type of morpheme that can only be contributed by the ML in bilingual CPs: "[A]ll system morphemes which have grammatical relations external to their head constituent (i.e. which participate in the sentence's thematic role grid) will come from the [ML]" (Myers-Scotton, 1993, p. 83). Under the 4-M model of morpheme classification described in the next section, this type of morpheme is defined more precisely.

3.1.2 The 4-M model

The 4-M model is a model of morpheme classification which was first introduced by Myers-Scotton and Jake (2000), taking up the content morpheme-system morpheme opposition introduced in the MLF model and developing it further by differentiating several types of system morphemes (Myers-Scotton, 2002, p. 16). The classification refers to the morphosyntactic roles of morphemes and links these surface roles to a model of language production (Levelt, 1989) considering how they are accessed in the speech production process (Myers-Scotton & Jake, 2017, p. 341). The differentiation of the different morpheme types is based on the mechanisms that activate them, assuming that "lemmas underlying different types of morphemes become salient at different levels of production" (Myers-Scotton & Jake, 2000, p. 1053). However, note that the notion of 'morpheme' is used in two different ways in the model. On the one hand, the term may refer to actual morphemes at the surface level. On the other hand, when referring to elements interacting in the abstract procedures that underlie language production, the term 'morpheme' is used metaphorically and refers to the lemmas that support the actual surface morphemes as abstract entries in the mental lexicon (Myers-Scotton, 2002, p. 17).

The 4-M model is not only an addition to the MLF model developed specifically for the analysis of classic codeswitching. Rather, the abstract distinctions underlying the 4-M model are claimed to reflect a universal organization of the mental lexicon (Myers-Scotton, 2002, p. 74; Myers-Scotton & Jake, 2000, p. 1054) that helps to explain various mono- and bilingual phenomena, such as codeswitching, speech errors, aphasia, and SLA (Myers-Scotton & Jake, 2000, p. 1053).

The model provides a four-way classification of morphemes. The primary distinction is that of content morphemes versus system morphemes. As already mentioned above, content morphemes convey semantic and pragmatic meaning (Myers-Scotton & Jake, 2000, p. 1054). Therefore, it is assumed that the lemmas underlying content morphemes are 'conceptually activated', i.e., they are activated by language-specific semantic and pragmatic feature bun-

dles that are selected at the conceptual level to convey speaker intentions (Myers-Scotton & Jake, 2000, p. 1058). Content morphemes either assign or receive thematic roles (Myers-Scotton & Jake, 2000, p. 1058). This feature distinguishes them from all types of system morphemes (Myers-Scotton & Jake, 2000, p. 1061). Content morphemes can occur independently of other elements in the clause. (Myers-Scotton & Jake, 2000, p. 1061). Prototypical examples for content morphemes are lexical verbs assigning thematic roles, nouns receiving thematic roles (Myers-Scotton & Jake, 2000, p. 1058), and adjectives and adverbs as they carry semantic meaning and modify the meanings of nominals and verbs (Myers-Scotton & Jake, 2017, p. 343). In contrast, system morphemes do not participate in the assignment of thematic roles (Myers-Scotton & Jake, 2000, p. 1058). They realize grammatical relations between constituents in a surface structure (Myers-Scotton & Jake, 2000, p. 1054). Prototypical system morphemes are function words and inflections (Myers-Scotton & Jake, 2000, pp. 1058–1059).

Furthermore, following the observation that "not all functional elements pattern alike" with regard to their distribution in actual utterances (Myers-Scotton, 2002, p. 71; Myers-Scotton & Jake, 2000, p. 1053), the 4-M model distinguishes three different types of system morpheme, some of which are activated at the conceptual level together with their content morpheme heads while others are structurally assigned at later stages of the production process (Myers-Scotton & Jake, 2000, p. 1053). Myers-Scotton and Jake use the terms of 'early system morphemes' for conceptually activated system morphemes opposed to 'late system morphemes' for structurally assigned system morphemes which only become salient at the level of the Formulator where syntactic structures are built (Myers-Scotton & Jake, 2000, pp. 1061–1062).[2]

The first type of system morphemes, the 'early system morphemes', occur together with content morphemes to express semantic and pragmatic features that are needed to further realize speaker intentions (Myers-Scotton & Jake, 2000, p. 1062). In contrast to content morphemes which are directly elected by semantic or pragmatic feature bundles from speaker intentions to map conceptual structure onto lemmas, early system morphemes are 'indirectly elected' by the content morphemes which "point to" them (Myers-Scotton & Jake, 2000, p. 1066). Particularly, they add specificity and transitivity readings to their content morpheme heads, expressing semantic and pragmatic concepts such as definiteness, plurality, completeness, and progressive (Myers-Scotton & Jake, 2017, p. 344) or *phi*-features such as person, number, and gender (Myers-Scotton, 2002, p. 75). Examples for early system morphemes are the English plural morpheme *–s* (Myers-Scotton & Jake, 2000, p. 1066) as well as the French determiners *le* and *la* 'the' selected by their head nouns to encode both definiteness and the *phi-features* of number and gender (Myers-Scotton, 2002, p. 81). Early system morphemes rely on their content morpheme heads for information about their form and can thus not occur independently, even though they can be free or bound depending on the way the language they occur in realizes morphosyntactic structures (Myers-Scotton & Jake, 2017, p. 344).

The second type of system morphemes, 'late system morphemes', differs from both content morphemes and early system morphemes in realizing grammatical information without

[2] Yet, they do make it clear that this terminology is only used as a metaphor for ease of reference and not intended to imply any statements about the relative time of access or claims about the ordering of access processes in language production (Myers-Scotton & Jake, 2017, p. 342).

expressing semantic meaning (Myers-Scotton & Jake, 2000, p. 1063). They are structurally assigned at the level of the Formulator to build larger constituents (Myers-Scotton & Jake, 2000, p. 1063). The 4-M model differentiates two sub-types of late system morphemes. Firstly, there are 'bridge late system morphemes' which satisfy grammatical configurations within a maximal projection and thus depend on their own maximal projection for their form (Myers-Scotton & Jake, 2000, p. 1064). Bridge late system morphemes integrate elements into larger constituents, e.g., joining together two units within a clause or uniting two clauses, and marking hierarchical relationships between the elements they connect (Myers-Scotton, 2002, p. 78; Myers-Scotton & Jake, 2017, p. 344). For instance, in English, the possessive preposition *of* and the possessive suffix -'*s* both act as bridge late system morphemes, connecting noun heads with their complements and expressing genitive without assigning theta roles as in *friend of Bora* or *Bora's friend* (Myers-Scotton, 2002, p. 79). Just like early system morphemes, bridge late system morphemes can be free or bound elements depending on how the language they occur in expresses the respective grammatical functions (Myers-Scotton & Jake, 2017, p. 345). Yet, bridge late system morphemes are not co-indexed at all (Myers-Scotton, 2002, p. 79). Secondly, 'outsider late system morphemes' as the second type of late system morpheme do not occur in the same constituent as the elements that activate them (Myers-Scotton & Jake, 2017, p. 345). Their form depends on grammatical requirements from outside the immediate maximal projection they occur in (Myers-Scotton & Jake, 2000, p. 1064). Outsider late system morphemes show co-indexical relationships between elements across different maximal projections (Myers-Scotton, 2002, p. 78), making relationships between elements in a clause more transparent and realizing the grammatical frame of a clause (Myers-Scotton & Jake, 2017, pp. 345–346). They are only accessed when larger constituents like CPs or IPs are constructed (Myers-Scotton, 2002, p. 76). Prototypical outsider late system morphemes are found in agreement morphology. In English, the subject-verb agreement marker *-s* is an outsider late system morpheme which depends on the subject and can therefore not be realized until the IP containing it is joined with a third-person singular DP[3] in the subject position (Myers-Scotton, 2002, p. 80).

Two comments have to be made concerning the 4-M model. Firstly, the classification of morpheme types is not bound to lexical categories. Rather, it may be observed that within a lexical category, some elements are identified as content morphemes and others as system morphemes depending on their morphosyntactic roles. This is the case for English pronouns: While personal pronouns may receive thematic roles and therefore be defined as content morphemes, expletive pronouns like *it* and *there* usually are inserted into CPs to ensure grammaticality without receiving thematic roles and, thus, act as bridge late system morphemes. English prepositions are another case: Some assign thematic roles, e.g., *for* in assigning the

[3]In earlier works, Myers-Scotton and Jake (2000, p. 1060) disclaim the DP hypothesis by Abney (1987) stating that the determiner is the functional head governing the noun phrase (NP). This changes in more recent works where they predominantly work with DPs as functional projections, even though not consistently (e.g., Myers-Scotton & Jake, 2017). The choice of either position does not affect the results of this thesis. For the sake of consistency and in agreement with the more recent developments by Myers-Scotton and Jake, the DP analysis is adopted here. See also Keller (2020, pp. 25–26) arguing for the inclusion of the DP as a functional projection in the investigation of codeswitching data.

role of BENEFACTIVE to *Jane* in *I made a cake for Jane*, and are thus characterized as content morphemes. Others, like the bridge late system morpheme possessive *of*, only realize hierarchical structure without showing thematic relationships (Myers-Scotton & Jake, 2017, pp. 343–344).

Secondly, the mapping of lexical category to morpheme type may vary cross-linguistically (Myers-Scotton & Jake, 2017, p. 343). For example, the English determiner *the* adds definiteness to the noun it occurs with and is categorized as an early system morpheme under the 4-M model. In contrast, German determiners do not only convey definiteness (and *phi*-features such as number and gender) but they also carry case inflections assigned by the verb outside their own maximal projection. Therefore, they can only be fully realized at the level of the Formulator where they receive the final information about their form. Thus, German determiners are not early but outsider late system morphemes (Myers-Scotton & Jake, 2017, p. 346). Note that German determiners represent a special case with regard to their classification under the 4-M model as the features of gender and number they express are early system morphemes while case is a late system morpheme. Hence, the lexemes unite different types of morpheme and are therefore called 'multimorphemic'. However, (Myers-Scotton, 2002, pp. 81–82) assumes that the late system morphemes take precedence in the classification because the final form of the surface morphemes does not become salient until the level of the Formulator.

The 4-M model adds precision to the MLF model as it helps to define the type of morpheme that is subject to the system-morpheme principle (Myers-Scotton, 2002, pp. 86–87). With the 4-M model, this type of morpheme can be labeled as outsider late system morpheme (Myers-Scotton, 2002, p. 87). While early system morphemes and bridge late system morphemes may come from either language in bilingual CPs, only the ML may contribute outsider late system morphemes in classic codeswitching (Myers-Scotton & Jake, 2017, p. 345). Moreover, Myers-Scotton claims that the classification of morpheme types applies universally. That is, it accounts for the differential distribution of surface morphemes in language-contact phenomena beyond classic codeswitching (Myers-Scotton, 2002, p. 85).

3.1.3 The Abstract Level model

The Abstract Level model was first outlined in Myers-Scotton and Jake (1995). It supports the MLF model by working out the details of how EL morphemes are checked for sufficient congruence with their ML counterparts before being integrated into the ML frame. Also, it helps to explain contact phenomena other than classic codeswitching by providing explanations for how abstract grammatical structure from different languages may be combined to build a composite ML (Myers-Scotton, 2002, p. 96). The Abstract Level model is closely linked to Levelt's (1989) model of speech production.

The central assumption in the Abstract Level model is that every lexical entry in the mental lexicon includes three levels of abstract lexical structure: 1) lexical-conceptual structure, 2) predicate-argument structure, and 3) morphological-realization patterns (Myers-Scotton & Jake, 1995).[4]

[4]The phonological information, which, according to Levelt (1989), also constitutes a part of information of a lexical entry, is not considered by Myers-Scotton and Jake (1995), as it does not play a relevant role in the grammatical encoding procedures the model focuses on.

The assumption underlying the postulation of a level of lexical-conceptual structure in lexical entries is that languages differ in how universally available semantic and pragmatic concepts are lexicalized (Myers-Scotton & Jake, 1995, p. 991; Talmy, 1985). For example, in English, the concepts of 'knowing sth.', 'knowing so.', and 'knowing to do sth.' are conflated in the verb *to know*. In French, in contrast, these meanings are expressed with different verbs: *connaître qc./qn.* meaning 'to know sth./so.' is used with animate and inanimate objects while *savoir faire qc.* is used to express the meaning of 'knowing to do sth.', e.g., *savoir lire* 'know to read' Myers-Scotton and Jake (1995, p. 987) assume that the 'lexical-conceptual structure' of abstract entries in the mental lexicon contains information about how these universally available semantic and pragmatic features are conflated into language-specific 'semantic/pragmatic feature bundles'.

At the level of 'predicate-argument structure', the lexical entry contains information about how thematic structure is mapped onto grammatical relations (Myers-Scotton, 2002, p. 19), e.g., concerning the assignment of thematic roles to the arguments of a verb and the realization of these in grammatical surface structures such as PPs or DPs (Myers-Scotton, 2002, p. 96; Myers-Scotton & Jake, 1995, p. 1001). For example, the predicate-argument structure underlying the verb *wait* contains the information that the thematic role of GOAL is assigned to its object which is realized as a PP in English, e.g., *I am waiting* [*for the bus* (GOAL)]$_{PP}$.

Finally, the level of 'morphological-realization patterns' includes information on how grammatical relations are realized in surface structure, e.g., regarding morpheme order and agreement morphology such as case and subject-verb agreement (Myers-Scotton, 2002, p. 19; Myers-Scotton & Jake, 2017, p. 348).

The Abstract Level model is relevant for the MLF model and explanations of classic codeswitching as it informs the concept of 'sufficient congruence' (Myers-Scotton, 2002, p. 19), assuming that a match of abstract lexical structure of EL and ML morphemes serves as a precondition for full morphological integration of EL morphemes into the ML frame in classic codeswitching (Myers-Scotton, 2002, p. 20). Crucially, sufficient congruence does not mean complete congruence (Myers-Scotton, 2002, p. 20) but rather that the features encoded in the EL content morpheme on the different levels of abstract structure must satisfy the requirements of the ML frame (Myers-Scotton, 2002, p. 97). Yet, Myers-Scotton admits that it is not clear what exactly constitutes sufficient congruence (Myers-Scotton, 2002, p. 20).

Moreover, the Abstract Level model also informs explanations for grammatical structure observable in language-contact phenomena other than classic codeswitching. More precisely, it helps to account for the structure of the abstract morphosyntactic frame structuring bilingual clauses in contact phenomena that demonstrate a composite ML, i.e., where the abstract structure underlying the surface utterances comes from more than one language (Myers-Scotton, 2002, p. 19). The central assumption here is that bilingual CPs can be projected by abstract lexical material from more than one language. This happens when speakers do not have full access to the desired ML, e.g., in situations of language shift, creole formation, or L2 acquisition (Myers-Scotton, 2002). In these cases, abstract material from lexical entries in one language can be split off and combined with abstract structure from another language, resulting in a composite. Thus, in composite codeswitching, the ML as the abstract grammatical frame is a

composite of abstract lexical structure from several sources rather than derived from only one of the participating languages (Myers-Scotton, 2002, p. 22). This process is seen as the major mechanism underlying the construction of interlanguage under the CML model of SLA.

3.1.4 Criticism of the Matrix Language Frame model

Over time, criticism of the MLF model has arisen. First of all, it has to be mentioned that the MLF model has continuously developed since the publication of its original version in 1993 and some of the early assumptions have been changed or discarded altogether. This section briefly addresses two main arguments that are frequently brought up.

For example, in the first version of the model (Myers-Scotton, 1993), it has not always been made clear that the system-morpheme principle does not apply to all system morphemes in general but only to those having grammatical relations external to their head, which has led to misunderstandings in subsequent work by other researchers (Myers-Scotton, 2002, p. 87). Also, in the 1993 version of the model, a quantitative morpheme count has been suggested for the determination of the ML in samples of bilingual speech, which has been dismissed from 1997 onwards (cf. the *Afterword* of the second edition of *Duelling Languages*, Myers-Scotton, 1997) The early version of the model has been criticized by several researchers (Bentahila, 1995; Meechan, 1995; Muysken & Rooij, 1995).

From 1997 onwards, two principles have been used to identify the ML of bilingual CPs (Myers-Scotton, 2002, p. 59). The first one is the system-morpheme principle introduced in section 3.1.1, claiming that outsider late system morphemes may only come from the ML in bilingual CPs. The second one is the 'morpheme-order principle' stating that morpheme order in bilingual CPs has to be that of the ML as well (Myers-Scotton, 1993, p. 83). The use of these two principles as tests for the status of the ML as well as the continuous evolution of the criteria for determining the ML have evoked methodological criticism (Boussofara-Omar, 2003; Poulisse, 1998). Particularly, it is often argued that the MLF model is circular because the ML is identified based on the distribution of morpheme types and morpheme order and then used to explain these same aspects: "However, the purely structural definition is somewhat circular if the matrix language thus determined is then invoked to explain the origin of system morphemes such as the verbal inflections and the complementizer." (Muysken, 2000, pp. 67–68).[5]

Myers-Scotton has rejected this criticism by clarifying that the terms contained in the morpheme-order principle and system-morpheme principle ('morpheme order' and 'system morphemes' which have grammatical relations external to their head constituent) are independent of the theoretical construct of the ML and can be defined objectively (Myers-Scotton, 2002, p. 59). In addition, despite the methodological criticism, research into the MLF model proves that its assumptions and predictions can be confirmed based on a broad range of empirical data (see section 3.3.1). With this criticism in mind, the MLF model and the CML model as its

[5]The criticism of the MLF model is more extensive than illustrated here, and also involves controversies about potential counter-examples to the principles stated under the MLF model. However, a full discussion is beyond the scope of this thesis. See Myers-Scotton (2010) for an exemplary overview of some objections brought up by other researchers and her responses to these.

extension are used and tested in this thesis. If their underlying assumptions can be verified, this will serve as further evidence for their universal applicability and relevance in modeling language-contact phenomena.

3.2 The Composite Matrix Language Model

3.2.1 Theoretical foundations and goals of the model

The goal of the CML model is to explain SLA as a language-contact phenomenon and to account for the structure of IL, focusing on how acquisition and transfer are connected (Jake, 1998, pp. 333–334). The model aims to describe how IL as a developing system projects a grammatical frame structuring surface constituents (Jake, 1998, p. 334) and to find out how transfer (i.e. L1 influence) is regulated in IL construction (Jake, 1998, p. 336), constraining what each participating linguistic system may contribute to the construction of IL (Jake, 1998, p. 333). Additionally, it strives to predict "what kinds of IL structures are possible, what underlies IL structures, and what direction IL development will take" (Jake, 1998, p. 335). A further claim is that the model generalizes to SLA data independently of the particular languages involved (Jake, 1998, p. 337).

The central argument underlying the extension of the MLF model to SLA is that interlanguage development can be viewed as an instance of language contact where the participating languages, namely L1 and TL, interact in the construction of interlanguage as a separate linguistic system of its own and in the projection of interlanguage surface utterances. It is assumed that in the process of SLA, learners intend to speak the TL, which is therefore labeled the 'intended ML'. Yet, due to the incompleteness of the acquisition process, the grammatical system of TL and the abstract lexical structure of TL lexical entries are not always fully available to speakers. As a consequence, gaps in the abstract lexical structure of TL lemmas may be filled by abstract material from the L1 acting as an EL. The resulting abstract grammatical frame is a composite of abstract lexical structure from L1 and TL, acting as the 'de facto' ML projecting the abstract grammatical frame underlying actual IL surface utterances (Jake, 1998). Thus, in L2 acquisition, contact takes place at an abstract level of linguistic structure preceding surface-level structures (Jake, 1998, p. 334).

3.2.2 Assumptions and predictions of the Composite Matrix Language model

3.2.2.1 The target-language principle and the complete-projection principle guiding second-language acquisition and transfer The first and most central assumption made in the CML model of SLA is that "second-language learners intend to speak the TL" (Jake, 1998, p. 341). This assumption is related to learner intentions and leads to the stipulation of two crucial principles guiding the construction of IL.

The first one is the 'target-language principle': "To the extent possible, construct the IL from TL lexical structure" (Jake, 1998, p. 341). It leads to the prediction that "[a]ll IL surface structures are projected by TL-based lexical items in the grammatical system underlying IL" (Jake, 1998, p. 342). This means that at the abstract level of language production, "real or

putative" TL lexical structure must be identifiable as underlying all IL lexical items (Jake, 1998, p. 342), even though the L1 may contribute abstract lexical structure in order to fill in gaps in incompletely acquired TL lexical items (see section 3.2.2.2 below). In other words, IL structures are not predicted to be projected by L1 abstract lexical structure exclusively. However, as we do not have direct access to the acquired TL-knowledge of the learner, we cannot determine with certainty what the actual grammar underlying a learner's production looks like.

Examples potentially violating the target-language principle at the abstract level are not considered by Jake (1998). Possible counter-examples could be loan translations of fixed phrases including verbs, such as proverbs or idioms, where not only the lexical-conceptual structure and predicate-argument structure of the L1 but also L1 morphological-realization patterns for word order, case, and agreement are employed in the construction of IL surface structures. Such examples might be most clearly detectable in the IL of learners whose L1 and TL differ typologically with regard to their lexicalization patterns, thematic role assignment and realization, word order, and morphological devices for case and agreement marking.

At the surface level of IL utterances, the target-language principle implies that only the TL may contribute surface forms in IL structures. Exceptions are cases of codeswitching occurring when the learner assumes that his interlocutor knows his L1 as well and false cognates[6] (Jake, 1998, p. 342).

The second principle driving IL development is the 'complete-projection principle': "To the extent possible, satisfy the requirements of the grammar of the matrix language through the specification of all requisite grammatical features of the entries in the mental lexicon (i.e. lemmas)" (Jake, 1998, p. 342). It is based on Chomsky's 'projection principle' stating that the properties of lexical elements in the mental lexicon must be represented by the syntactic structure projected by them (Chomsky, 1993, p. 29). The complete-projection principle transfers this to the context of SLA, implying that the three levels of abstract lexical information that are relevant for the construction of grammatical IL surface structures need to be "filled out" as completely as possible, e.g., regarding the assignment of theta roles, case, and *phi*-features (Jake, 1998, p. 342). In cases where the learners' access to TL lemmas is incomplete, the principle allows for transfer of abstract lexical structure from L1 lemmas. However, what is transferred from the L1 are not wholesale lexemes but only those parts of abstract lexical information of L1 lexical entries that are necessary to completely specify the missing information ('gaps') in the incompletely acquired TL lexical entries (Jake, 1998, p. 346).

Additionally, it is predicted that transfer of abstract lexical information from L1 lemmas is only possible if it is construed as sufficiently congruent with acquired TL-based structure, i.e., L1 transfer is subject to congruence checking with and therefore determined by TL (Jake, 1998, p. 341). Yet, this condition is problematic with regard to its verification in several ways:

[6]Unfortunately, Jake does not define explicitly what she means by false cognates. In the example she uses to illustrate learners resorting to false cognates in their IL production, it becomes clear that she is referring to words that have a similar form in L1 and TL (Jake, 1998, p. 375). These words may be used by learners hoping that they represent actual cognates with both similar form and meaning in their L1 and the TL, hence fulfilling the learner's intention to express themselves in the TL. Depending on the actual lexical items selected by the learners, the term of false cognates as used by Jake may thus refer to true cognates, false cognates and even false friends.

Firstly, 'acquired TL-based structure' is not necessarily the same as TL structure conforming to a certain standard or norm, and we do not know what exactly an individual learner's acquired TL structure looks like. Secondly, the concept of sufficient congruence is not clearly defined. Thirdly, IL structures do not tell us what has not been transferred, so we cannot determine when transfer has been blocked.

3.2.2.2 Interlanguage as a projection of composite lexical structure from the native and the target language

Referring to the Abstract Level model, the mechanism of IL construction under the CML model is based on the assumption that IL lexical structure is a composite "made up of TL and L1 lexical structure whose abstract levels of sub-structures are split and recombined" (Jake, 1998, p. 347). More precisely, this prediction implies that the requirement to specify abstract information in TL lexical entries as completely as possible may be accomplished in two ways: either by the insertion of parts of abstract lexical structure from lexical entries in the learner's L1, representing transfer, or by the extension of information contained in TL lexical entries that are already available to the learner, representing overextension of TL material (Jake, 1998, p. 346).

In the process of filling gaps in the abstract lexical structure in TL lexical entries, the L1 can provide abstract material on all three levels of lexical structure. Firstly, when the lexical-conceptual structure of a lexical item in TL is not completely available, learners may resort to the lexical-conceptual structure of a (sufficiently) equivalent item in their L1, leading to inappropriate lexical choices of TL lexemes in IL production. For example, a learner of English with the L1 Chinese who has not yet learned that the meanings 'to open sth.' and 'to turn sth. on' are lexicalized in different verbs in English may transfer lexical-conceptual structure from his L1 where these meanings are conflated in one and the same verb, as in *Open air condition* 'Turn on the air condition' (L1 Chinese/L2 English; Wei, 1995; as cited in Wei, 2009, p. 12), resulting in the election of the English verb *open* to express the meaning 'turn on' (Wei, 2009, pp. 12–13).

Transfer of abstract material from the L1 may also happen at the level of predicate-argument structure, as in *Sometimes I watch TV or listen radio* (L1 Japanese/L2 English; Wei, 1994; as cited in Jake, 1998, p. 349). In the TL English, the THEME of *listen* has to be expressed as a PP (*to the radio*). Due to incomplete acquisition of the predicate-argument structure of this verb, the learner relies on predicate-argument structure from the L1 Japanese where the THEME of *listen* is expressed as an internal object (Jake, 1998, p. 349), resulting in the realization of the object *radio* as an internal object in IL as well.

Finally, the L1 may also influence TL structures on the level of morphological-realization patterns. For example, the word order of IL utterances may be influenced by the L1 as in *In Japan student English junior high school start* (L1 Japanese/L2 English; Wei, 1994; as cited in Jake, 1998, p. 352). Here, the sentence-final position of the verb *start* is transferred from the learner's L1 Japanese because the corresponding information is not yet acquired in the TL (Jake, 1998, p. 351).

3.2.2.3 Order of morpheme acquisition in second-language acquisition Apart from the Abstract Level model, also the 4-M model plays a role in IL development (Jake, 1998, p. 343), leading to the assumption that the order of morpheme acquisition in SLA follows their salience in the abstract process of language production. According to this prediction, content morphemes and early system morphemes, which are conceptually activated, are acquired earlier than late system morphemes, which are structurally assigned at the level of the Formulator (Jake, 1998, p. 343). This prediction has been tested and confirmed empirically (Wei, 1996a, 2000a, 2000b, 2003; see also section 3.3.2) and is not investigated further in this thesis.

3.2.2.4 Constraints on transfer from the native language in the construction of interlanguage Furthermore, the content morpheme-system morpheme opposition and the system-morpheme principle related to it have an impact on the extent to which the influence from the L1 is limited in the construction of IL, constraining how L1 lexical material can fill gaps in the lexical structure of ML items at an abstract level (Jake, 1998, p. 362).

On the one hand, abstract material from the L1 may not fill any gap in the ML, but "only those gaps that the learner's grammar specifies as being projected by content morphemes" (Jake, 1998, p. 362). This includes early system morphemes which are indirectly elected by content morphemes (Jake, 1998, p. 363). In contrast, the L1 may not fill in any gap in the ML that is projected by outsider late system morphemes. This prediction is, in turn, difficult to test empirically as we do not know whether a specific TL morpheme is specified as a content morpheme in the individual learner's IL grammar.

On the other hand, "not just any lexical structure projected by the L1 can fill an ML gap; only lexical structure projected by L1 content morphemes can fill in a gap in the ML" (Jake, 1998, p. 362). Again, this includes early system morphemes. However, abstract lexical structure underlying L1 late system morphemes cannot be transferred into the ML underlying IL surface structures.

Unfortunately, Jake (1998) does not exemplify what potential counter-examples to the constraints on L1 influence in the CML model might look like. These might be most easily imaginable if the L1 and the TL are typologically different concerning their morphosyntactic properties. For example, while in English, the subject-verb agreement marker *-s* is attached to the verb as a suffix, Swahili has subject markers attached to the verb as prefixes (Mpiranya, 2014, p. 13). This information about the position of these morphemes in relation to the verb they are attached to represents abstract lexical structure at the level of morphological-realization patterns underlying outsider late system morphemes which may not be transferred in SLA. Thus, in the acquisition of the L2 English with the L1 Swahili, the system-morpheme principle would disallow the transfer of the positional realization of the Swahili late system morpheme subject-verb agreement marker as a prefix instead of a suffix as required by the L2 English.

3.2.2.5 The unrestricted role of the target language in filling gaps in interlanguage Due to the restricted role of L1 in the process of filling gaps in TL lexical items in the construction of IL, some gaps in the ML, especially those projected by late system morphemes, cannot be filled by L1 material (Jake, 1998, p. 363). These may either remain unfilled, reflecting the phenomenon

of fossilization in interlanguage development, or they may eventually be filled by the insertion of TL lexical structure (Jake, 1998, p. 363). Because the TL acts as the intended ML in SLA, the CML model predicts that TL lexical structure is not constrained in how it fills gaps in the composite syntactic frame underlying IL surface utterances (Jake, 1998, p. 363). Under this prediction, any type of TL lexical structure is allowed to fill system morpheme gaps, including both abstract lexical structure and entire lexical items in the form of surface morphemes (Jake, 1998, p. 364). However, it should be noted that it cannot be proven that open gaps in IL are instances of fossilization and not the result of, e.g., motivational aspects, learner attitudes, or (possibly intentionally employed) compromise strategies like avoidance.

Crucially, even though the insertion of TL elements satisfies the complete-projection principle, it does not necessarily result in target-like IL structures (Jake, 1998, p. 369). For example, learners who have not yet acquired the inflectional paradigm for verbal inflections for person and number may overextend the use of forms that they already know, e.g., the suffix for third-person singular, and insert these into late system morpheme gaps where another suffix, e.g., expressing first-person singular, would be required (Jake, 1998, p. 363). Only when new TL material becomes available to the learners, new projections will replace former IL structures and IL will become more target-like (Jake, 1998, p. 363). Hence, Jake (1998, p. 341) supposes that the acquisition process is lexically driven and that the target-language principle drives the continuous development of the grammatical system underlying IL (Jake, 1998, p. 341).

3.2.3 Summary: The Composite Matrix Language model

The CML model of SLA (Jake, 1998) is based on the framework of the MLF model (Myers-Scotton, 1993, 1997, 2002). According to the matrix language-embedded language opposition stated under the MLF model, only one of the languages participating in bilingual speech, namely the ML, provides the abstract grammatical frame underlying surface utterances. The EL, in contrast, is restricted to contributing conceptually activated elements that may be inserted into the ML frame. The 4-M model (Myers-Scotton & Jake, 2000) specifies that outsider late system morphemes may only come from the ML in bilingual speech. The Abstract Level model (Myers-Scotton & Jake, 1995) makes predictions for language contact situations where the desired ML is not completely available to the speakers. In these cases, a composite ML may arise: Different levels of abstract lexical structure from different languages may be split off and recombined, resulting in a composite of abstract lexical structure projecting the grammatical frame underlying surface structures.

Under the CML model, it is assumed that contact in SLA takes place at an abstract level of lexical structure prior to surface structures. The model makes several predictions based on the principles of language contact stated in the framework of the MLF model. Firstly, based on the target-language principle, the TL is identified as the intended ML in second-language acquisition. As a consequence, the CML model predicts that abstract lexical structure underlying TL surface morphemes must be identifiable as underlying all surface structures in IL and that only the TL may contribute surface forms in IL structures.

Secondly, the complete-projection principle states that all three levels of abstract lexical

information that are relevant for the construction of grammatical IL surface structures need to be "filled out" as completely as possible. However, L2 learners do not have complete access to the abstract lexical structure underlying TL due to incomplete acquisition. Hence, parts of abstract lexical structure from L1 surface elements may be transferred to fill in these gaps in TL. The result is IL, a composite of abstract lexical structure from L1 and TL. In second-language acquisition, IL represents the de facto ML which projects the abstract grammatical frame underlying actual IL surface structures. Based on the complete-projection principle, the CML model predicts that IL surface structures must be explicable as projections of composite abstract lexical structure from L1 and TL.

Finally, as the L1 acts as an EL in second-language acquisition, it is constrained in what it may contribute to fill in gaps in the composite ML. In accordance with the predictions made under the 4-M model, abstract lexical structure from the L1 may not fill in gaps that are projected for outsider late system morphemes in the composite ML. In addition, abstract lexical structure underlying L1 outsider late system morphemes may not be transferred to the composite ML. However, the TL as the intended ML is not constrained in any way in filling gaps in the composite ML. More specifically, both abstract lexical structure and surface morphemes from TL may be inserted into gaps for content morphemes, early system morphemes, and late system morphemes in IL, representing the overextension of TL material.

The following chapter gives an overview of previous research into the explanatory power of the principles underlying the MLF model and the predictions of the CML model, forming the basis for the development of the research questions guiding this thesis in chapter 3.4.

3.3 Literature review

The first section of this chapter presents research illustrating the relevance of the principles stated under the MLF model in explaining the structural outcomes of various language contact phenomena. The second section focuses on studies providing evidence for the assumptions made in the CML model, also addressing how this thesis may contribute further evidence for the claim that the model generalizes to SLA data in general.

3.3.1 Application of the Matrix Language Frame model to language-contact phenomena

Based on the principles established under the MLF model, the 4-M model, and the Abstract Level model, Myers-Scotton (1993, 1998, 2002) has stated concrete predictions for how these principles apply in language-contact situations other than classic codeswitching, especially those involving the development of a composite ML. In particular, she predicts the possible outcomes of language contact in situations involving L1 attrition, convergence, and creole formation. These predictions have been tested by several researchers thereafter, mostly providing evidence for the extended version of the MLF model.

According to Myers-Scotton (1998, 2002), the attrition of an L1 necessarily involves convergence. In this view, attrition refers to the loss of a speaker's first language due to convergence of his L1 to the L2 (Myers-Scotton, 2002, p. 231). Convergence, in contrast, is viewed as the main mechanism creating structural changes in L1 attrition (Myers-Scotton, 2002, p. 242). As

a covert type of bilingual speech, it involves the splitting and recombining of abstract lexical structure from more than one language, resulting in a restructuring of the abstract grammatical frame underlying bilingual speech, i.e., a composite ML, while only one of the participating languages contributes surface morphemes (Myers-Scotton, 2002, pp. 164–165).

Based on the Abstract Level model and the 4-M model and their implications for language production, Myers-Scotton has made several predictions for the susceptibility to change of the different levels of abstract lexical structure and for the distribution of different morpheme types in convergence and (L1) attrition. In a hierarchy of relative susceptibility to change through convergence, the lexical-conceptual structure is assumed to be the most susceptible to change as it is closest to speaker intentions and therefore easily accessed and altered (Myers-Scotton, 2002, p. 196). The level of predicate-argument structure is assumed to be the least susceptible to change as the information contained in the subcategorization frame of a verb is less flexible (Myers-Scotton, 2002, p. 205). Concerning the different morpheme types, the prediction is that content morphemes are the first to be 'lost' or replaced in the L1 and the first type of L2 morpheme entering the L1 because of their easy accessibility at the conceptual level (Myers-Scotton, 2002, p. 206). Early system morphemes are more susceptible to change than late system morphemes which are only activated at the level of the Formulator and therefore least accessible (Myers-Scotton, 2002, p. 207).

Several studies on structural change at the level of the speech community provide empirical evidence for the predictions made by Myers-Scotton for convergence and L1 attrition. For instance, Fuller (1996) investigates contact-induced language change from the 1940s to the 1990s in Pennsylvania German spoken by Amish and Mennonite communities where English gains importance as a language of everyday communication and church services in the community (Fuller, 2000, p. 49). Fuller argues that changes in Pennsylvania German can be accounted for by convergence to English at the level of abstract lexical structure resulting in the emergence of a composite ML underlying Pennsylvania German surface structures. She demonstrates how abstract structure on the levels of lexical-conceptual structure and morphological-realization patterns from English is introduced into the abstract morphosyntactic frame underlying surface utterances in Pennsylvania German, leading to the overextension of progressive aspect marking and word order changes (Fuller, 1996).

In a follow-up study, Fuller (2000) examines the behavior of different morpheme types in the process of convergence of Pennsylvania German to English. She demonstrates that early system morphemes such as past participles and the plural marker *-s* are the first type of system morpheme to be influenced by the L2 English due to their conceptual salience while there is little evidence of any influence from English regarding late system morphemes in the Pennsylvania German data.

In a similar study, Jones (2018) examines the predictions by Myers-Scotton for attrition and tests whether these also apply for language loss at the level of the community in the case of Jersey Norman French (Jèrriais), a variety of French spoken by the elder generations in Jersey as an L1 which is in extensive contact with and slowly replaced by English as the main everyday language (Jones, 2018, pp. 404–405). Jones shows that at the level of lexical-conceptual structure convergence with the L2 English leads to the abandonment of semantic distinctions

in Jèrriais, e.g., between 'yes' in affirmative contexts (*oui*) and in contradictory contexts (*si*) (Jones, 2018, pp. 408–409). At the level of morphological-realization patterns, Jones cites examples such as word order changes concerning adjectives and their head nouns (Jones, 2018, pp. 411–412). She even provides evidence for convergence at the level of predicate-argument structure, e.g., the omission of reflexive pronouns in Jèrriais due to English influence (Jones, 2018, p. 415). In addition, Jones provides quantitative evidence confirming that content morphemes are most susceptible to change by demonstrating that Jèrriais contains many English-origin nouns and verbs that are morphologically integrated into the language (Jones, 2018, pp. 417–418). She also shows that while the L2 English affects Jèrriais early system morphemes such as plural markers and definite articles (Jones, 2018, p. 419), Jèrriais late system morphemes are very rare to be replaced influenced by English late system morphemes (Jones, 2018, pp. 421–423).

In sum, the studies cited here confirm the predictions made by Myers-Scotton (1998, 2002) for the structural outcomes of language attrition and convergence at the level of the community, providing evidence for the relative susceptibility to change of different levels of abstract lexical structure and of different morpheme types.

Apart from work examining the applicability of the principles underlying the MLF model in convergence at the level of the speech community, other studies shed light on convergence at the level of individual speakers. For instance, Fuller and Lehnert (2000) investigate how NPs are structured in German-English codeswitching (where German is the ML) with regard to the roles of the participating languages in gender assignment and the projection of the article in adult speakers. They demonstrate that classic codeswitching and convergence may co-occur under specific sociolinguistic circumstances (Fuller & Lehnert, 2000, p. 418). In a group of native German speakers living in the U.S., they find both patterns of classic codeswitching in gender assignment operating according to German (=ML) strategies and, simultaneously, patterns of convergence and composite codeswitching in the projection or non-projection of articles which sometimes followed English (= EL) rules (Fuller & Lehnert, 2000). These findings underline that the MLF model can not only account for cases of classic or composite codeswitching but also for the dynamics between these phenomena.

Furthermore, several studies focus on convergence and attrition in child bilingualism, emphasizing that the principles stated under the MLF model can account for both overt and covert instances of bilingual speech, which may be interrelated and co-occur in individual speakers.

For example, Bolonyai (1998) retraces alternating patterns of classic codeswitching, convergence, and composite codeswitching as the structural outcomes of bilingual language acquisition of Hungarian (L1) and English (L2) in an L2-dominant environment. Her study follows the development of a Hungarian child (ages between 3;7 and 4;10). The data demonstrate how the ML changes in classic codeswitching according to the sociolinguistic circumstances. In the instances of convergence, which are sometimes accompanied by overt codeswitching, influence of English on Hungarian is visible at the level of lexical-conceptual structure, mostly resulting in inappropriate lexical choices, and at the level of morphological-realization patterns, leading to non-target-like word order or case realizations. These patterns peak after a stay in Hungary, indicating that the intention to communicate in Hungarian is thwarted by

incomplete access to the grammatical system underlying it, which leads to the emergence of composite structures due to English influence (Bolonyai, 1998). The study demonstrates that changes in the dominance relationship between the L1 and the L2 and the resulting structural outcomes are mainly determined by socio- and psycholinguistic factors.

In subsequent studies on the bilingual acquisition of Hungarian and English, Bolonyai (2000, 2002) focuses on the acquisition of Hungarian case markers by immigrant children living in the U.S. with the L1 Hungarian as their home language and English as the dominant language of the larger society. Hungarian case inflections show differential error rates depending on inter-categorial variation concerning their status as early system morphemes versus late system morphemes. The Hungarian case-marking system includes both lexically-determined and structurally assigned cases. The structural cases involve nominative and accusative which are structurally assigned as late system morphemes, while the lexically determined cases comprise oblique cases such as locatives which are specified in the lexical entry of the subcategorization frame of the verb and therefore categorized as early system morphemes (Bolonyai, 2000, p. 91). Bolonyai reports that early system morpheme case markers show lower error rates than late system morpheme case markers compared to the total number of non-target-like case markers in the data (Bolonyai, 2000, p. 97). This result contradicts the prediction that in language attrition, late system morphemes are least susceptible to change. However, the error rates should not be compared to the total number of incorrectly produced case markers but to the respective frequencies of occurrence of structural and lexical case markers in the data. This has been considered in Bolonyai (2002) who notes an asymmetry in the actual probability of occurrence of the accusative versus oblique cases with the accusative being significantly more frequent than oblique cases in mono- and bilingual speakers of Hungarian. When the actual frequencies of occurrence of early system morpheme case markers versus late system morpheme case markers are taken into account, structural case markers show lower error rates compared to the actual number of occurrence than oblique case endings, which is in line with the prediction that late system morphemes are the least susceptible to change.

Finally, Schmitt (2000) witnesses the co-occurrence of classic and composite codeswitching in a heritage language context, observing the bilingual production of Russian immigrant children in the U.S. She demonstrates that gaps in the intended ML Russian may be filled by classic codeswitching ranging from the insertion of English content morphemes with full morphosyntactic integration into the Russian-based frame to the insertion of EL constituents structured according to English rules (Schmitt, 2000, pp. 17–18). However, she cites data indicating that the high amount of gaps in the abstract lexical structure of L1 morphemes may also trigger abstract lexical structure from English to participate in the projection of the abstract frame underlying bilingual speech, leading to convergence and eventually resulting in the production of covert codeswitching where all surface morphemes are provided by the ML Russian but the abstract frame is a composite (Schmitt, 2000, pp. 19–22).

The studies concerning language convergence and attrition at the level of the individual cited here demonstrate that the transition between overt and covert forms of bilingual speech is not clear-cut but fluent. Still, the principles underlying the MLF model can predict and explain structures in both contexts of classic and composite codeswitching as well as conver-

gence.

Further evidence for the applicability of the extended version of the MLF model to language-contact phenomena other than classic codeswitching comes from creole formation. In an article on the case of Berbice Dutch, a creole that developed as a result of language contact between Dutch and Eastern Ijɔ during the 17th and 18th centuries in the Berbice colony (present-day Guyana), Gross (2000) demonstrates that the principles from the Abstract Level model and the 4-M model can also account for the structural processes underlying creole formation.

According to Myers-Scotton (2002, pp. 273–274), the ML or the abstract grammatical frame underlying surface structures in a creole is a composite projected by abstract lexical structure from the languages involved in the contact situation, i.e., the substrate(s) as the language(s) spoken by the enslaved people, and the superstrate as the language of the slave owners (Myers-Scotton, 2002, p. 274). These languages serve as two targets in the process of creole formation (Myers-Scotton, 2002, p. 273). As the superstrate is available to the enslaved people only to a limited extent, it serves as a target for lexical elements, namely content morphemes and early system morphemes, which may be inserted into the frame as content morphemes and early system morphemes or reanalyzed to fill in late system morpheme gaps in the ML. Hence, the superstrate may also be called the 'lexifier'. In contrast, the substrate(s) are fully available to the enslaved people and thus psycholinguistically serve as a target for the abstract morphosyntactic frame of the developing creole (Gross, 2000, p. 62; Myers-Scotton, 2002, p. 273).

In his study, Gross (2000) provides evidence that late system morphemes from the incompletely available superstrate do not occur in the creole grammar. However, pointing out the importance of the perceptual salience of morpheme types, he claims that there are two exceptions where late system morphemes from the superstrate may occur in the creole, both involving reanalysis (Gross, 2000, p. 66). Firstly, late system morphemes that have multiple status in the superstrate, meaning that may occur as early system morphemes or late system morphemes depending on the context, may be reanalyzed as early system morphemes and enter the creole in late system morpheme contexts as a result of this reanalysis. Secondly, late system morphemes may enter the creole as unanalyzed forms that are bound to content morphemes (see Gross, 2000, p. 66 for examples from Guyanais Creole French and Guyanese Creole English). The study by Gross (2000) constitutes a further piece of evidence for the relevance of the principles from the MLF model in governing various language-contact phenomena.

In sum, the studies discussed in this section confirm the explanatory power of the principles underlying the MLF model and the CML model, justifying their application for the explanation of language contact data and SLA despite the methodological criticism mentioned in section 3.1.4.

3.3.2 Evidence for the Composite Matrix Language model

In the studies cited above, it has become evident that the transitions between different contact phenomena are fluid. This has also been claimed to be the case for language contact and SLA. For instance, Odlin & Yu (2016) state that the difference between transfer and codeswitching is "subtle or even non-existent" (Odlin & Yu, 2016, p. 2). One study positioning itself within

research claiming that there are parallels between classic codeswitching and interlanguage development is Fuller (1999). Claiming that all bilingual discourse is structured by the same abstract principles, she aims to compare bilingual speech in these two language-contact situations to show that similar structures arise which can be explained by the MLF model (Fuller, 1999, p. 540).

However, there are some problematic aspects regarding the selected data and methodology in the study that need to be addressed here. First of all, the interlanguage data analyzed in the study comes from an elder learner of English who has two L1s, namely Spanish and German (Fuller, 1999, p. 535). This makes it difficult to determine which language is responsible for specific structural features in the English interlanguage data. Moreover, one of the L1s, German, which appears to be the main source for transfer, has not been used regularly in the subject's adult life (Fuller, 1999, p. 535) and might, therefore, be undergoing attrition processes leading to structural changes in the L1. These might, in turn, have consequences for the lexical information that may be transferred to the L2 English. This is problematic for the analysis of the IL data as there is not enough information about the individual grammar of German influencing IL structures. In addition, the codeswitching data compared to the IL structures come from persons who only have German and English as their languages (Fuller, 1999, p. 536), which makes the overall comparability of these data with the interlanguage data questionable.

On the methodological side, it has to be mentioned that the analysis and discussion of the data are affected by misconceptions in the application of the MLF model and its related models. To identify the ML in her codeswitching data, Fuller (1999, p. 541) relies on quantitative morpheme counts, a criterion that has been abandoned by Myers-Scotton from 1997 onwards. Furthermore, Fuller does not differentiate between early and late system morphemes, a distinction which is crucial for the discussion because according to the system-morpheme principle, only late outsider morphemes have to come from the ML in bilingual speech.

Finally, the interlanguage data cited in the analysis majorly involve overt codeswitching between German and English and sometimes even Spanish. However, there is no explicit differentiation of whether the ML is a composite or comes from only one language in these IL data. They are mostly analyzed as instances of classic codeswitching without considering that the abstract frame underlying these utterances might be a composite of abstract lexical structure from both L1 and L2.

All in all, the data cited by Fuller (1999) cannot really be used as evidence for or against the MLF model as we do not know enough about the grammatical system underlying the subject's L1. However, despite the methodological issues, the study by Fuller (1999) demonstrates that interlanguage development may be seen as one of many language-contact phenomena which can be conceived of as a continuum of bilingual speech ranging from overt to more covert effects, depending on the speakers' degrees of bilingual proficiency and of access to at least one of the participating languages, rather than separated phenomena.

Apart from research confirming the applicability of the MLF model to different language-contact phenomena and SLA in general, the specific predictions of the CML model have been tested in subsequent work. In particular, an unpublished corpus of learner data from Chinese

and Japanese learners of English (Wei, 1995) has played an important role in supporting the CML model of SLA. It has originally been collected for Wei's – also unpublished – doctoral thesis (Wei, 1996b) where he investigates distributions and accuracy rates of different morpheme types in the acquisition of English as an L2 and relates them to different levels of election at an abstract level in language production.

The corpus underlying Wei's research has been described in Wei (2000a, 2000b, 2015). It comprises interlanguage data from 60 learners of English in total, of which 30 are native speakers of Chinese and 30 have Japanese as their L1. The participants are international students at a university in the U.S. and their families residing in Columbia, South Carolina. Following a classification of learner stages used by the *European Science Foundation* (Klein, Dietrich, & Noyau, 1993; Klein & Perdue, 1993), the learners were divided into pre-basic, basic, and beyond-basic level for each L1 background according to their proficiency in the L2 English. Data collection involved question-based interviews to approximate natural conversation as well as picture-description tasks eliciting "descriptions related to 'existence, location, possession, condition, etc.' and (…) descriptions involving 'ongoing, completed, or future activities' " (Wei, 2000a, p. 34). The data contained in the corpus or sub-sets of it have served as the empirical basis for Wei's subsequent work on the CML model.

First of all, Wei (1996a, 2000a, 2000b, 2003) has used data from that corpus to provide statistical evidence for the order of acquisition of the different morpheme types predicted by Jake (1998). In an early article, Wei (1996a) compares the distributions and error rates of different morpheme types in basic and pre-basic learners of English and observes that content morphemes and semantically transparent system morphemes (which would be called early system morphemes under the 4-M model) are acquired earlier than syntactically relevant system morphemes without semantic content (late system morphemes under the 4-M model) (Wei, 1996a, pp. 425–426), even though he does not relate these findings to how these types of morphemes are accessed in language production.

In his subsequent work on the distribution and accuracy of different types of morphemes in Chinese and Japanese learners of English, Wei tests the specific predictions established by Jake (1998) for the order of acquisition of morpheme types in SLA. His studies on the implicational hierarchy of morpheme acquisition all follow the same rationale. He focuses on the acquisition of morphemes in particular lexical categories, such as pronouns or determiners, which show variation regarding the categorization of their members as content morphemes, early system morphemes, or late system morphemes under the 4-M model. By comparing accuracy distributions of the respective morphemes in learners at the pre-basic, basic, and beyond-basic levels, he draws conclusions regarding which morpheme types in a lexical category are learned early or late in the acquisition process, and accounts for these observations by relating them to how the different morpheme types are accessed in language production.

For instance, Wei (2000a) investigates the lexical categories of determiners, pronouns, and prepositions in pre-basic and basic learners. In the class of determiners, he compares accuracy distributions of early system morpheme possessives like *my, your, their* and early system morpheme demonstratives like *this* and *that* against the definite article *the* which is classified as a late system morpheme occurring in constructions where it is activated at the level of

the Formulator to build syntactic structure, as in *in the hospital* (Wei, 2000a, p. 33). Concerning pronouns, Wei compares accuracy distributions of content morpheme personal pronouns and freestanding demonstratives as in *this is the right answer* against the late system morpheme dummy pronouns *it* and *there* (Wei, 2000a, p. 34). In the category of prepositions, content morpheme prepositions assigning thematic roles like *for* are compared to the late system morpheme possessive preposition *of* (Wei, 2000a, p. 34). For all three categories, Wei (2000a) provides statistical evidence that morphemes that are activated earlier in the process of language production also demonstrate lower error rates, indicating that the level of activation affects learning difficulty (Wei, 2000a, p. 40).

In Wei (2000b), two further lexical categories are included in the investigation of the order of acquisition of morpheme types. The first category that is added includes inflections of lexical verbs. Here, the early system morpheme progressive *-ing* and past participle *-ed/-en* are compared to the late system morpheme subject-verb agreement marker *-s* and past tense *-ed* (Wei, 2000b, pp. 115–116). In the second category, namely verbs under INFL, accuracy distributions of content morpheme modals such as *can, should, will* participating in the assignment of thematic roles are compared to late system morphemes auxiliaries *have, do, be* which are activated by syntactic procedures to satisfy grammaticality requirements without assigning thematic roles or adding semantic meaning to the message (Wei, 2000b, pp. 117–118). The statistical analysis of error rates in pre-basic and basic learners provides further evidence that morpheme types that are conceptually salient and therefore accessed 'early' in language production are acquired earlier than those that are structurally assigned at the level of the Formulator (Wei, 2000b).

Wei (2003) mainly repeats the findings reported in Wei (2000a) and Wei (2000b), again examining the order of acquisition of different morpheme types in the categories of verb inflections, determiners, and pronouns. All in all, it can be concluded that the order of morpheme acquisition proposed by Jake (1998) can be confirmed for the learners in Wei's (1995) corpus, even though these findings are not representative of all Chinese and Japanese learners of English or for SLA in general. Further research is needed to validate the predictions of the CML model for the order of morpheme acquisition in SLA to a more general extent.

In addition to the study of the acquisition of different morpheme types, examples from the corpus by Wei (1995) have been cited by Jake (1998) as well as by Wei (2009, 2015, 2018) and Wei and Liu (2017) to illustrate the predictions made in the CML model for how abstract lexical structure from the L1 and the TL may be split off and recombined in the projection of a composite grammatical frame underlying IL surface structures. In these articles, the assumptions underlying the CML model of SLA are recapitulated and explained in more detail. Examples from the corpus are provided for transfer of abstract lexical structure from the L1s Chinese and Japanese to English on the three levels of 1) lexical-conceptual structure, 2) predicate-argument structure, and 3) morphological-realization patterns.

One further study by Wei (2006) applies the assumptions and predictions underlying the CML model to third-language acquisition. The data investigated in the study come from four learners: two native speakers of Chinese with an advanced level in the L2 English learning German as an L3 and two native speakers of Japanese with the L2 English learning Chinese

as an L3 (Wei, 2006, p. 96). Based on the interlanguage data from L3 acquisition, Wei (2006) demonstrates that the principles underlying the CML model also apply to third-language acquisition: When the (intended) ML is not fully available, abstract lexical structure from either the L1 or a previously, possibly incompletely, acquired L2 is split up and recombined with abstract material from the TL to fill in gaps in the ML, resulting in a composite ML. Specifically, with his data indicating that the learners heavily relied on their L2 English in the construction of their L3 interlanguages independently of the relative similarity of their L2 and L3, he concludes that in third-language acquisition, previously learned interlanguages have a particularly strong influence on the development of the interlanguage underlying L3 production.

Apart from Wei's work, there is one study that relies on a different sample of learners. Liu (2015) applies the CML model to data from Japanese learners of English. She demonstrates that also in the case of Japanese learners of English, transfer takes place at all three levels of abstract lexical structure, resulting in the construction of a composite ML underlying utterances in the learners' IL as predicted by Jake (1998). Her analysis, however, is selective as only the most common errors have been considered (Liu, 2015, p. 237).

In sum, it has to be stated that while the studies mentioned above help to spell out the predictions of the CML model in a more detailed way, their contribution of empirical support for the model remains limited for two reasons. Firstly, the studies are not necessarily representative as they do not reflect exhaustive analyses but rather illustrative selections of learner data, leaving it unclear whether the data could also exhibit counter-examples. Secondly, one claim made in the model is that the predictions for the development and structure of IL generalize to any L1-TL pair. However, the data illustrating the predictions are restricted to different pairings of Chinese, Japanese, and English as L1s and L2s, which is by far not enough to confirm the universal applicability of the model. By analyzing data from English learners of the L2 French, this thesis seeks to contribute to testing whether the CML model can also account for IL structures with a different language pair in order to provide further evidence for the general relevance of the principles hypothesized to structure language contact phenomena. Additionally, the goal is to not only selectively present illustrative data supporting the models but to take into account IL structures that seem to be problematic with regard to the predictions made, aiming to provide a more comprehensive analysis of the data. In the following chapter, the concrete research questions guiding the analysis of the IL data of English learners of the L2 French are formulated, focusing on whether the predictions made under the CML model can be confirmed.

3.4 Research questions

This thesis aims to investigate the claim of the CML model that it can explain IL surface structures as projections of abstract lexical structure from L1 and TL independently of which languages are involved. Specifically, it applies the predictions of the model to IL data from English learners of French. Thus, the superordinate research question of this thesis is: Can the CML model of SLA by Jake (1998) account for IL surface structures observed in learners with the L1 English and the L2 French?

Firstly, if the CML model can be shown to account for the IL structures found in the learner data examined in this thesis, this will support the claim that the CML model generalizes to SLA data independently of the specific L1-TL pair involved. Secondly, the question is also a test of the universal principles that are assumed to govern the grammatical outcomes of language-contact phenomena as stated under the MLF model (Myers-Scotton, 1993, 2002). If the predictions of the CML model can be confirmed, this will also represent further evidence for the general relevance of the principles of language contact stated under the MLF model.

For the analysis, the superordinate research question is divided into four more specific research questions targeting the specific predictions made under the CML model:

1. Are the IL surface structures observed in the learner data based on TL lexical items?

2. Can the IL surface structures observed in the data be explained as projections of composite lexical structure from the L1 and TL?

3. Do the restrictions for the limited influence of the L1 hold?

4. Are there IL structures that cannot be accounted for within the framework of the CML model and, if yes, what are their implications for the predictions of the model?

The first research question (RQ 1) relates to the target-language principle. At an abstract level, we test whether IL surface structures are always projected by TL-based lexical structure. However, this is difficult to prove because if the abstract structure underlying lexical items in the participating languages is similar, it cannot be determined with certainty where it comes from. At the surface level, RQ 1 tests whether only TL contributes surface forms. The findings regarding RQ 1 are discussed in chapter 5.1.

The second question (RQ 2) refers to the complete-projection principle predicting that both the L1 and TL may fill in gaps in IL to specify the requirements of the TL lexical entries projecting IL surface structures as completely as possible. Here, it is investigated whether all IL structures in the data can be explained as based on composite lexical structure from L1, via transfer of abstract lexical structure on three levels, and from TL, via overextension and insertion of TL material at the surface and abstract levels. The findings regarding RQ 2 are discussed in chapter 5.2 presenting evidence for the predictions of the CML model and chapter 5.3 considering IL structures that cannot be straightforwardly explained as projections of composite lexical structure from L1 and TL.

The third question (RQ 3) tests the constraints on the influence from the L1 as implied by the system-morpheme principle, i.e., that transfer from the L1 is restricted to abstract lexical structure underlying L1 content morphemes or early system morphemes being inserted into gaps in IL projected by TL content morphemes or early system morphemes. Again, there should be no wholesale transfer of complete L1 morphemes but only transfer of parts of abstract lexical structure from L1 to fill in gaps in incompletely acquired TL material. If the constraints can be confirmed based on the learner data, this will support the relevance of psycholinguistic concepts such as language dominance and differential activation of morphemes

in language production underlying the MLF model in predicting and explaining the structural outcomes of language contact. The findings regarding RQ 3 are discussed in chapter 5.2 illustrating IL structures conform to system-morpheme principle and chapter 5.3 discussing potential counter-examples.

Finally, the fourth question (RQ 4) is aimed at IL structures that cannot be explained as the result of the interaction of L1 and TL as predicted under the CML model. These structures are considered in chapter 5.3 with regard to their implications for the principles underlying the CML model. The following chapter presents the data the analysis is based upon.

4 Method

4.1 Data

The data analyzed in this study come from the *Newcastle Corpus* (Myles & Mitchell, 2013b) originating from the research project *The Structure of French Interlanguage: A corpus-based study* funded by the Arts & Humanities Research Council (Myles & Mitchell, 2013c). It is available via the *French Learner Language Oral Corpora* (FLLOC) database, resulting from a series of research projects of the same name. The FLLOC projects, which started in the mid-nineties and ended in 2013, were funded by the UK Economic and Social Research Council, the Arts & Humanities Research Council, and the British Academy (Myles & Mitchell, 2013a). The FLLOC database contains several learner corpora that document and promote research into the development of interlanguage in the acquisition of French as an L2 in the classroom context in the UK with learner ages ranging from five- to seven-year-old primary school pupils to university students in their mid-twenties (Myles & Mitchell, 2013a). The *Newcastle Corpus* contributes to this goal by providing spoken data from learners with an intermediate level of French in their final years of college attendance, enabling the analysis of interlanguage in different domains including syntax, morphology, and discourse (Myles & Mitchell, 2013c).

4.1.1 Participants

The *Newcastle Corpus* includes spoken interlanguage data from 30 British students from four publicly-funded sixth-form colleges who were recorded in year 12 and again in year 13. At the beginning of the study, the participants were aged 16-17 and in their sixth year of learning French. The recordings reflect an intermediate L2 level (Myles & Mitchell, 2013d).[7]

Apart from information about learner age and sex, information concerning the learners' language biographies was collected as well. Not surprisingly, some of the participants do not only speak English and French but also have one or several other previously learned languages at their disposal, e.g., Spanish, Latin, German, and Dutch, which they master to different degrees, ranging from a limited knowledge of rarely used L2s learned at school to native languages other than English used on an every-day basis. As previous research in second- and

[7] For an overview of the learners' sex, age, grades in French, and of the tasks available per participant, see table A1 in the appendix.

third-language acquisition has demonstrated, previously learned interlanguages may play a major role as a source for transfer in the acquisition of an additional foreign language (Bardel & Falk, 2007; Wei, 2006). Especially, L2s that are genetically or typologically related to the TL, or perceived as such by the learners, are preferably chosen as sources for transfer, even though the level of proficiency and recency of use also represent deciding factors (Falk & Bardel, 2010). Consequently, the existence of other previously learned languages apart from the L1 English makes it impossible to analyze the data with regard to the roles of L1 and TL as it cannot be determined with certainty whether a particular structure is caused by influence from the L1 or by a previously – possibly incompletely – acquired L2. Therefore, learners with knowledge of languages other than English and French were excluded from the analysis to avoid interference. As a result, data from ten students were analyzed in this thesis.

4.1.2 Tasks

The participants had to perform five different oral tasks on a one-to-one basis with a researcher and one pair discussion task together with another participant (Myles & Mitchell, 2013e). Two oral tasks included story-retelling: In the "Loch Ness" task (LT), learners had to re-tell a story that had been narrated by the interviewer before. They were allowed to use a picture book and a vocabulary list. In the "Modern Times" task (MT), learners were shown an excerpt from a silent film by Charlie Chaplin and then asked to retell the scene with the help of a vocabulary list (Myles & Mitchell, 2013e).

In addition, there were three elicitation tasks. The "Interrogatives" task (QT) eliciting interrogatives was based on a picture including four characters, some of which were omitted in the learners' version of the drawing. Learners had to ask questions about the missing characters, e.g., concerning their appearance and their activities, in order to draw them. The second elicitation task, "Negatives and Adverbs" (NA), was based on cartoons showing persons doing different activities. In this task, learners had to describe in one simple sentence what the persons shown in the cartoon were doing. When the pictures were crossed out, they had to use negation in their description and when an adverb was written below the picture, it had to be included in the sentence. The third elicitation task, the "Photos" task (PT), was based on three pictures of young people in a summer or a winter scene. First, learners had to ask questions about the persons shown in the pictures to gain as much information as possible about them. In a second step, the interviewers asked the learners about their own summer or winter holidays using future and past tense (Myles & Mitchell, 2013e).

Finally, in the pair discussion task (PD), the two participating learners were allowed to choose one out of four topics. In this task, they were asked to debate and subsequently rate a list of suggested measures that comprise potential solutions to the problem at hand (Myles & Mitchell, 2013e).

In the corpus, separate sound files are available per task and speaker, making a total of about one hour of spoken language per participant and recording round (Myles & Mitchell, 2013e). In addition, morphologically tagged transcripts that conform to the CHILDES conventions (MacWhinney, 2000) are provided (Myles & Mitchell, 2013f). In some cases, not all transcrip-

tions are available. For the analysis, only data from the first recording round were considered. All tasks were included in the analysis so that the discussion is based on an extensive source of data involving different linguistic structures.

4.1.3 The subcorpus analyzed in this study

The subcorpus created for this study includes 2,138 tokens of CPs from 10 learners. Table 1 shows the distribution of L1, L2, mixed, and interlanguage CPs in the subcorpus.

Table 1: Number of CPs in the subcorpus

L1	L2	Mixed	IL	Total
270	615	105	1148	2138

The subcorpus contains 270 (=12.63%) monolingual CPs in the participants' L1 English. These are briefly discussed in chapter 5.1 with regard to their implications for the target-language principle stated under the CML model. However, as they do not represent instances of transfer at an abstract level, they are not further analyzed in this study.

A total of 615 (=28.77%) CPs in the subcorpus can be labeled as target-like, i.e., representing CPs in the participants' L2 French and conforming to TL norms. While these might also be seen as instances of IL, they do not provide information about potential transfer processes at the abstract level underlying these surface structures. Therefore, these CPs are not analyzed further in this study.

Apart from CPs in the participants' L1 or L2, the subcorpus contains 105 (=4.91%) mixed CPs, i.e., CPs comprised of surface elements from both L1 and TL. The implications of the occurrence of L1 surface elements in the mixed CPs for the target-language principle are addressed in chapter 5.1. As these CPs may also show instances of transfer of abstract lexical structure, they are included in the qualitative analysis.

Finally, there are 1,148 (53.70%) CPs representing IL structures in the subcorpus. More specifically, these are the CPs that involve at least one instance of non-target-like structures.[8] Together with the mixed CPs, they form the basis for the analysis. In sum, a total of 1.253 CPs has been analyzed in this study.

4.2 Analysis

To put the predictions from the CML model to the test, the data were analyzed comprehensively and all non-target-like structures were documented. Grammaticality of IL structures was examined with reference to the grammar of French by Batchelor and Chebli-Saadi (2011). In order to verify the meanings of morphemes, the *Oxford Dictionary of English* (Stevenson,

[8]Note that the number of CPs containing IL structures only gives an orientation regarding the quantity of CPs that have been analyzed. However, it does not represent an indicator for the actual proportion of non-target-like elements in the data as these would have to be counted on the morpheme level.

2010) and the *Petit Robert* (Robert & Rey, 2014) were used. However, several comments have to be made concerning what was analyzed as non-target-like structures.

Firstly, the French language entails several registers, some of which allow for grammatical structures that deviate from the standard. IL structures deviating from standard French but conforming to one of these registers were not counted as non-target-like structures because they represent stylistic/pragmatic choices rather than morphosyntactic errors. For example, the formation of *yes/no*-questions in French involves subject-verb inversion or insertion of the question marker *est-ce que* in formal registers while in more familiar and colloquial registers, the formation of the same kind of question is acceptable with direct order and a rising intonation pattern to mark the interrogative. Thus, IL structures involving *yes/no*-questions marked by intonation were analyzed as target-like because they conform to the colloquial register in French.

Secondly, many structures involving tense depend on the context of the narration. To avoid undue interpretation, the use of present instead of past or future tenses was only counted as non-target-like when either the question asked by the interviewer or the immediate context of the learner's utterance (e.g., the use of temporal indicators such as *l'été prochain* 'next summer') required the respective tense to be used.

Finally, IL structures often involve repetitions with different versions of the sentence. False starts that were repaired were discarded as instances of performance and thus not counted as non-target-like.

The non-target-like structures detected in the data were first classified descriptively. First of all, it was checked whether these IL structures are based on TL material at the abstract and surface levels. At the surface level, all IL structures that include only TL surface morphemes were classified as TL-based. Structures involving surface morphemes from the L1 were classified as not exclusively TL-based and were further categorized as complete CPs in the learners' L1, instances of codeswitching, or loan blends (see chapter 5.1). At the abstract level, IL structures that are explicable as projected by abstract lexical structure underlying lexical items from TL were classified as TL-based. This includes IL structures whose abstract grammatical frames are analyzable as projected by TL abstract lexical structure on the levels of 1) lexical-conceptual structure, 2) predicate-argument structure, and 3) morphological-realization patterns, even though gaps in the abstract lexical structure underlying IL may have been filled with L1 material (see chapter 5.2). In contrast, IL structures whose underlying morphosyntactic frames are not explicable as projected by abstract lexical structure from TL were classified as non-TL-based. This applies to IL structures whose grammatical frame is only explicable as projected by abstract lexical structure from the L1, or to IL structures that are not explicable as projections of composite lexical structure from L1 and TL (see chapter 5.3).

In a second step, IL structures classified as TL-based were categorized into instances of L1 influence on the respective levels of 1) lexical-conceptual structure, 2) predicate-argument structure, and 3) morphological-realization patterns, or instances of TL material inserted into gaps in the IL. Furthermore, the instances of L1 influence were checked for whether the constraints stipulated in the CML model, i.e., that L1 structure may only come from content morphemes and early system morphemes and be inserted into gaps in IL content morphemes and

early system morphemes, hold. Finally, the non-target-like structures that were ambiguous or did not fit into any of these categories were further examined. The results of the analysis and examples are reported and discussed in the next chapter.

5 Results and discussion

In this chapter, the results of the analysis are presented. The discussion follows the order of the research questions posed in chapter 3.4. Chapter 5.1 deals with the question whether all IL surface structures are based on the TL at the abstract and surface levels (RQ 1). In chapter 5.2, it is demonstrated that non-target-like IL structures can be explained as projections of composite lexical structure from L1 and TL (RQ 2). Firstly, section 5.2.1 presents instances of IL surface structure representing transfer from the L1 on three levels of abstract lexical structure. The examples in the respective sections all conform to the constraints on L1 influence resulting from the system-morpheme principle (RQ 3). Secondly, section 5.2.2 covers instances of insertion of TL material into IL gaps. Thirdly, the occurrence of open gaps in IL as instances of fossilization are addressed in section 5.2.3. Finally, chapter 5.3 considers IL structures that cannot be straightforwardly accounted for under the CML model (RQ 4). This includes ambiguous cases concerning the morpheme type of the elements involved and cases of uncertainty with regard to what kind of structure the learners intended to produce. Additionally, instances of IL where neither L1 or TL abstract lexical structure nor a composite of abstract structure from both languages seems to be the source of IL structures are examined in this chapter. Furthermore, cases where the constraints on transfer of abstract lexical structure from the L1 seem to be violated are presented.

5.1 The target language as the basis of interlanguage surface structures

As outlined in section 3.2.2.1, the CML model includes the target-language principle as a central principle guiding the development of interlanguage, stating that learners seek to use the TL exclusively and to construct IL from TL structure as much as possible. The prediction resulting from this principle is that, at the abstract level, TL structure must be identifiable as underlying IL structures while at the surface level, all morphemes should come from the TL. Exceptions are cases of codeswitching when learners know that their interlocutors also understand their L1, or false cognates (Jake, 1998, p. 342).

Concerning the abstract level of lexical structure, it can be confirmed that the TL is the basis for the construction of all IL surface utterances, that is, the L1 does only contribute parts of abstract lexical structure from L1 morphemes to fill in gaps in the TL. In the data, no IL structures were found where only L1 abstract lexical structure on all three levels was the basis for the projection of IL surface structures. However, this finding could be attributed to the fact that English and French, the languages participating in the construction of IL in the data, are typologically similar at the levels of lexical-conceptual structure, predicate-argument structure, and, even if to a lesser extent, at the level of morphological-realization patterns. This makes it difficult to determine whether specific patterns come from L1 or TL. Thus, the lack

of counter-examples could also be due to the low chance of occurrence in this specific L1-TL pair.

At the surface level, the large majority of the IL structures contained in the data only include TL morphemes. Yet, a few exceptions were detected, as shown in (1) and (2). On the one hand, knowing that the interviewers also speak and understand English, the participants resort to codeswitching to achieve communicative goals. Example (1) demonstrates that this strategy may result in the production of complete CPs in English, e.g., to ask task-related questions (1a), to express uncertainty about vocabulary (1b), and, less frequently, to complete task-internal utterances (1c).

(1) a. DO I HAVE TO SAY IT?[9] (19 LT)[10]
b. I DON'T KNOW THE WORD A BLAZER A COAT (08 QT)
c. BECAUSE HE FELT SORRY FOR HER (…) HE PITIED HER (08 MT)

On the other hand, intra-CP or single-word switches occur. These switches are more frequent than complete CPs in English and mostly concern content morphemes, such as nouns (2a), verbs (2b), adverbs (2c), and discourse markers (2d) and are probably due to gaps in the learners' TL vocabulary. The English items are mostly inserted as bare forms without morphosyntactic integration, which could be explained by the limited knowledge of TL morphology.

(2) a. *une* SLIDE *ou une* SWING *ou juste une* TREE OR SOMETHING (08 QT)
a slide or a swing or just a tree or something
b. *et l' agent de police* SAY *non* (30 MT)
and the policeman say no
c. *oui* MAYBE (08 MT)
yes maybe
d. *non elle a les* (…)[11] LIKE *les cheveux brunes ou blondes ou* (08 QT)
no she has the like the hair brown or blond or?

Note that insufficient congruence does not seem to be a fitting explanation in (2) for several reasons. The switched nouns in (2a) represent concrete entities whose abstract lexical structure should be relatively equal to that of their French equivalents concerning their conceptual information and morphological-realization patterns. The same applies to the verb in (2b), especially with regard to the level of predicate-argument structure which often causes congruence issues with verbs in general, and the affirmative adverb *oui* in (2c). While the discourse marker in (2d) might be the switched element that does not find a conceptually congruent equivalent in the TL, it is not an element that needs morphosyntactic integration in

[9] In the examples from the *Newcastle Corpus* cited in the text, English morphemes are in CAPITALS; French morphemes are in *italics*.

[10] For every IL structure cited from the corpus, a reference is given in parentheses indicating the participant number, the task where the structure has been produced, and the recording round: e.g., 19 LT stands for participant number 19, Loch Ness task; 20 PD stands for participant number 20, pair discussion task.

[11] For reasons of readability, hesitations, false starts, repetitions, and reparations have been eliminated from the examples.

the TL. Thus, it cannot be deduced that incongruence is at work here at any level except from pragmatic meaning in the first place.

Apart from codeswitching, the data contain instances of loan blends as illustrated in (3). In (3a) and (3b), French verb inflections have been attached to English verb stem surface morphemes. In (3c), a French verb stem is the base for the derivation of an agentive noun using the English derivational affix *-er*.

(3) a. *elle* **CREAT**-*ait un monstre* (20 LT)
she create-ed a monster
'she created a monster'
[TL: elle a **cré**é un monstre][12]

b. *ils* **ESCAP**-*és* (23 MT)
they escape-ed
'they escaped'
[TL: *ils se sont* ***échap****pés*]

c. *il n' est pas une chant*-**ER** (20 NA)
he NEG is not a sing-er
'he is not a singer'
[TL: *il n'est pas un chant****eur***]

Loan blends are not mentioned by Jake (1998, p. 342) as possible exceptions to the prediction that TL must underlie all IL surface structures. Nonetheless, we can assume that they fall into the same category as codeswitching and false cognates because they represent cases where learners try to construct IL from TL material as much as possible: They resort to their L1 because they perceive it as typologically close to the TL French and know that their interlocutors know both languages as well but still rely on TL-based morphemes as much as possible. Note that the structures presented in (3) do not represent transfer of abstract lexical structure. Rather, they can be classified as instances of classic codeswitching where morphemes from the EL English are inserted into the ML frame and morphosyntactically integrated as completely as possible. This confirms that different language-contact phenomena such as codeswitching and interlanguage construction are related and may co-occur in the bilingual speech of individuals.

However, apart from cases of codeswitching and loan blends which are allowed as exceptions to the TLP, it can be stated that the IL structures found in the corpus conform to the

[12]The examples citing non-target-like structures are assembled as follows: The first line gives the IL structure as produced by the learner and the reference. The relevant non-target-like element is marked in **bold**. In the second line, a literal word-to-word translation is provided and, if relevant for the respective example, morphosyntactically annotated. The third line includes a figurative translation of the intended IL structure in English. In the fourth line, the correct TL version of the IL utterance is given and the relevant corrected element is marked in **bold**.

prediction that surface elements in IL always come from the TL. How TL and L1 abstract lexical structures interact in the construction of IL are discussed in the following chapter.

5.2 Interlanguage surface structures as projections of composite lexical structure from the native and the target language

The present chapter focuses on examples of IL structures found in the learner data that illustrate and therefore confirm the prediction that the composite grammatical frame underlying IL surface utterances results from a mechanism where information on different sub-levels of abstract lexical structure from the L1 or different types of TL elements may be inserted into gaps in the abstract lexical structure of lexical items in the IL mental lexicon (RQ 2). In the following, instances of L1 influence on three levels of abstract lexical structure and examples of insertion of TL elements into IL gaps are presented. Also, the examples cited here all conform to the constraints on L1 influence predicting that L1 influence may only come from conceptually activated morphemes in L1 and only be inserted into conceptually activated morphemes in IL (RQ 3).

5.2.1 Native language abstract lexical structure filling gaps in interlanguage

5.2.1.1 Native language lexical-conceptual structure in interlanguage As universal semantic and pragmatic features that learners may want to express are lexicalized differently in different languages, the transfer of lexical-conceptual structure from L1 may result in inappropriate lexical choices, as demonstrated in (4).

In (4a), the learner has not yet acquired the information that, in French, the meaning 'to leave' is lexicalized in different verbs: The verb *partir* expresses 'to leave, to depart' whereas the verb *quitter qc.* or *sortir de qc.* express 'to leave sth., to get out of sth.'. In contrast, the English verb *to leave (sth.)* conflates both the meaning 'to depart' and 'to get out of something', e.g., *They leave* versus *They leave the house.* Due to incomplete acquisition of the respective TL items, the learner resorts to their[13] L1 where one verb can be employed for both uses and uses the verb *partir* to express 'to get out of' in IL. In (4b), the learner has not yet acquired that French differentiates between the verbs *savoir* 'to know sth., to know how to do sth.' and *connaître* 'to know so.'. They thus transfer the lexicalization pattern from their L1 English where one verb (*to know sth./so.*) can be used in both situations, resulting in the election of the inappropriate verb in IL.

(4) a. *ils* ***partent*** *la* *maison* (19 LT)
they leave the house
'they leave the house'
[TL: *ils* ***quittent*** *la maison* / *ils* ***sortent de*** *la maison*]

[13] The sex and gender of the participants are not relevant for the analysis. Therefore, when referring to individual participants, singular *they* is used as a generic third-person pronoun in accordance with the *American Psychological Association*'s guidelines for biased-free language for gender (American Psychological Association, 2020).

b. *parce que* (…) *elle* (…) *n' a pas* ***savoir*** *le homme* (28 MT)
because she NEG has not know the man
'because she didn't know the man'
[TL: *parce qu'elle ne* ***connaissait*** *pas l'homme*]

5.2.1.2 Native language predicate-argument structure in interlanguage In some cases, the lexical-conceptual structure of verbs in IL may be intact but the predicate-argument structure is not completely available to the learner yet. Consequently, the transfer of predicate-argument structure from the learner's L1 to fill in gaps in incompletely acquired TL verbs may lead to different outcomes in IL.

On the one hand, it can result in IL structures where the assignment of thematic roles is intact but the realization of the object is not target-like, as in (5). Here, the thematic role for the object is assigned correctly, i.e., *le bus* is the GOAL of *attendre*. However, the object is realized as a PP according to the predicate-argument structure underlying the English verb *to wait for sth.* instead of French predicate-argument structure which requires the GOAL of *attendre* to be realized as a direct object DP.

(5) *il n' attend pas* ***pour*** *le bus* (24 NA)
he NEG waits not for the bus
'he is not waiting for the bus'
[TL: *il n'attend pas* ***Ø*** *le bus*]

On the other hand, L1 predicate-argument structure in IL may also result in incorrectly assigned thematic roles. This phenomenon was not detected in the data. However, this does not necessarily imply that the incorrect assignment of thematic roles due to L1 influence does not occur in SLA. Other researchers have confirmed that the L1 does, in some cases, influence the assignment of thematic roles in IL, especially with languages that are typologically different (Wei, 2009, 2018; Wei & Liu, 2017). Potential reasons for the absence of examples in the data examined include the fact that English and French are closely related languages which belong to the Indo-European language family and have undergone extended periods of contact in the past. Thus, the occurrence of differences concerning the assignment of thematic roles by verbs may be limited in number. In addition, learners tend to avoid L2 structures that are different from or non-existent in their L1. This has, amongst others, been shown to be the case for passive constructions, infinitive complements, relative clauses, and phrasal verbs (Dagut & Laufer, 1985; Kleinmann, 1977; Laufer & Eliasson, 1993; Schachter, 1974). Consequently, it also has to be considered here that learners may have avoided the transfer of predicate-argument structure in cases where role assignment differs between both languages due to (perceived) insufficient congruence of the respective verbs from their L1 and TL, even though this would be difficult to prove.

5.2.1.3 Native language morphological-realization patterns in interlanguage Finally, even if both the level of lexical-conceptual structure and the level of predicate-argument structure are intact, the L1 may still fill in gaps at the level of morphological-realization patterns, as illustrated in (6).

(6) *la personne* ***rarement téléphoner*** (23 NA)
the person rarely telephone
'the person rarely telephones'
[TL: *la personne* ***téléphone rarement***]

In this example, the word order of the verb and the adverb is not target-like. Here, word order is seen as an abstract early system morpheme as it underlies content morphemes (Myers-Scotton, 2002, p. 175). In French, the adverb of manner follows the verb it modifies. In English, in contrast, the adverb can either precede or follow the verb it modifies. As the learner has not yet acquired the respective morphological-realization patterns in French, they transfer the morphological-realization patterns from their L1 English into their IL production.

In sum, the examples discussed in section 5.2.1 have illustrated that L1 abstract lexical structure on three levels is transferred to IL to fill in gaps, confirming the complete-projection principle stipulated under the 4-M model stating that gaps in the abstract lexical structure underlying IL surface structure may be filled by parts of abstract lexical structure from the L1 to specify the missing information.

5.2.2 Insertion of elements from the target language into interlanguage gaps

Apart from transfer of abstract lexical structure from the L1, gaps in IL may also be filled by the insertion of TL material to satisfy the complete-projection principle. In contrast to the L1, the TL is not constrained in filling gaps in IL (Jake, 1998, p. 363). In the following, examples are given for surface lexemes and abstract lexical structure from the TL filling gaps in the learners' IL.

5.2.2.1 Target language content morphemes in interlanguage gaps Learners may use content morphemes from TL to fill in gaps projected by IL for which the appropriate morphemes are not yet available. Structures where French content morphemes have been inserted into gaps for content morphemes in IL are shown in (7).

In (7a), the noun *fils* 'son' is inserted into a gap for the content morpheme noun *garçon* 'boy' in IL which has assumably not been acquired yet by the learner. In (7b), the TL verb *faire* 'to do sth., to make sth.' substitutes the verb *fumer* 'to smoke'. This is a strategy that has occurred repeatedly with different learners who did not seem to know the verbs necessary to express the specific concepts they had in mind. In these cases, verbs with a high semantic intension, i.e., referring to well-defined activities, were replaced by the verb *faire* which has a high semantic extension and can, therefore, be used to circumscribe verbs referring to more specific actions that have not been acquired yet by the learners.

(7) a. *puis il a donné une cigare à une* (…) *petit **fils*** (27 NA)
then he has given a cigar to a little son
'then he gave a cigar to a little boy'
[TL: *puis il a donné une cigare à un petit **garçon***]

b. *il ne **fait de cigarette pas*** (25 NA)
he NEG does of cigarette not
'he does not smoke'
[TL: *il ne **fume** pas*]

5.2.2.2 Target language early system morphemes in interlanguage gaps Similarly to content morphemes, early system morphemes from TL may also be inserted into gaps in IL when the appropriate morphemes are not yet sufficiently available to the learners. Example (8) shows that the definiteness and indefiniteness in plural articles posed problems to the learners.

(8) *ils ont posé **les** question-s à la famille* (24 LT)
they have asked the.DEF.PL question-PL to the family
'they asked the family many questions'
[TL: *ils ont posé **des** questions à la famille*]

In French, the definite plural article *les* is used with countable nouns and entities whose number is exactly determined while the indefinite plural article *des* is required for uncountable nouns or nouns whose number is uncertain. English, in contrast, does not differentiate between a definite and an indefinite plural article. As the opposition in the TL is not yet completely available to the learner, the wrong form, i.e., the definite plural article, is inserted into the early system morpheme gap projected by the content morpheme *questions* 'questions' even though the number of questions is not exactly defined and thus the indefinite plural article would be required.

In addition to definiteness, French articles express *phi*-features, i.e., they include information about the grammatical gender and number of the nouns that elect them. This applies to definite as well as indefinite articles.[14] Throughout the data, there were many instances where inappropriate forms of the articles occur. The examples in (9) illustrate this phenomenon.

(9) a. *le* grand-mère (21 LT)
the.M grandmother
[TL: ***la** grand-mère*]

une homme (20 QT)
a.F man
[TL: ***un** homme*]

b. *le maison* (19 NA)
the.M house
[TL: ***la** maison*]

une bateau (21 LT)
a.F boat
[TL: ***un** bateau*]

[14]For an overview of the paradigms of French articles, see table A2 in the appendix.

c. *le trois enfants* (08 LT)
the.SG three children
[TL: ***les*** *trois enfants*]

mon grand-parents (27 PT)
my.SG grand-parents
[TL: ***mes*** *grand-parents*]

The examples in (9a) show inappropriate choices regarding the grammatical gender of the definite and indefinite article in cases where the grammatical gender of these articles is determined by the natural gender of the referents of the nouns they are accompanying. In the examples in (9b), grammatical gender has been assigned incorrectly with nouns that do not have natural gender. In these cases, morpho-phonological strategies could have given hints on the grammatical gender of the nouns (e.g., nouns ending in *-on* often have feminine gender while nouns ending in *-eau* mostly are masculine). However, it is uncertain whether the learners have learned these strategies. Finally, in (9c), there are two instances where the number of the nouns has not been respected. The examples cited here also illustrate that not only articles but also early system morpheme possessive adjectives (in this case, *mon* 'my.SG.M') may be subject to substitution by other forms of the respective TL early system morphemes.

In all these cases, we may argue that the different forms of the early system morpheme determiners have not been completely acquired yet and that therefore other forms are inserted into the gaps projected by the content morpheme nouns.

Verb inflections are another category that was subject to insertion of TL forms in the corpus. Some verb inflections, e.g., those for subject-verb-agreement, are categorized as outsider late system morphemes under the 4-M model. Others, such as participles and infinitive markers, are classified as early system morphemes that are indirectly elected by the verb they occur with to express aspect and do not look outside their own maximal projection for information on their form (Myers-Scotton, 2002, p. 95; Myers-Scotton & Jake, 2000, p. 1074; see also Wei, 2000b, p. 115 for a more detailed explanation). Examples for the insertion of TL early system morpheme verb inflections into gaps for these early system morphemes are presented in (10).[15] In (10a), a gap for the irregular past participle suffix *-ert* for the verb *découvrir* has been filled with the regular form of that suffix *-é* which represents a stereotypical case of overgeneralization. In (10b), the irregular past participle suffix *-u* has been substituted by the infinitive suffix of the verb. Again, both cases seem to be due to the incomplete acquisition of the respective participle forms.

(10) a. *la police* (…) *a découvr-**é*** *que* (…) *il a menti* (21 MT)
the police has discover-PPTCP that he has lied
'the police discovered that he lied'
[TL: *la police a découv**ert** qu'il a menti*]

[15] The examples given here focus on irregular verbs because in regular French verbs, the past participle ending *-é* [e] is homophonous and thus not distinguishable from the infinitive ending *-er* [e] in spoken language. However, this does not intend to imply that TL overgeneralization only occurs in irregular verbs — it may happen with regular verbs as well but this is impossible to prove with the TL French on the basis of spoken data alone.

b. *j' ai* (…) *recev-**oir*** (…) *beaucoup de* *cadeaux* (08 PT)
I have receive-INF many partART presents
'I received many presents'
[TL: *j'ai re**çu** beaucoup de cadeaux*]

Apart from determiners and verb inflections, early system morpheme prepositions have also been overextended and inserted into gaps for other early system morpheme prepositions that have not yet been available for the respective learners. This has happened with the prepositions *de* and *à* introducing the direct object of the verb *jouer* 'to play' in (11).

(11) *elle joue* ***du*** *football?* (08 QT)
she plays at.the football
'is she playing football?'
[TL: *elle joue **au** football*?]

In French, the complement of *jouer* may either be introduced by *à* when the noun refers to a specified game, e.g., *jouer au cache-cache* 'to play hide-and-seek' or by *de* when the noun refers to an instrument, e.g., *jouer du piano* 'to play the piano'. According to Myers-Scotton and Jake (2000, p. 1067), this type of preposition is classified as early system morpheme.[16] Thus, in (11), the masculine form of the early system morpheme preposition *de* is inserted into a gap where the verb *jouer* requires the masculine form of the preposition *à* to be used, representing a further case of overextension of TL morphemes to fill in gaps in IL.

Finally, in a few cases, early system morphemes from TL have also been inserted into gaps projected for TL late system morphemes. Example (12) illustrates structures where the early system morpheme definite article substitutes the late system morpheme partitive article. In French, a form of the partitive article *de* is required with uncountable nouns and indicates a part of a whole.[17] As a grammatical morpheme that satisfies the requirements within the maximal projection it occurs in and whose form depends on information within its own maximal projection, the French partitive article is categorized as a bridge late system morpheme (Myers-Scotton & Jake, 2000, p. 1064).

[16] Note that the prepositions *à* and *de* change their form when they contract with the masculine and plural form of the definite article: *à + le > au*, *à + les > aux*, *de + le > du, de + les > des.* In a description of the Spanish preposition *a* showing a similar contraction with the definite article, Myers-Scotton and Jake (2009, p. 349) do not assume that this process may change the status of the morpheme from conceptually activated to structurally assigned. As the prepositions *à* and *de* still only express conceptual information in their contracted forms and do not look outside their immediate maximal projection for information for their form, the view of Myers-Scotton and Jake (2009) is adopted in this discussion.

[17] The partitive article combines with the definite article and changes its form when it contracts with the masculine or plural article: *de + le > du*, *de + les > des.* In some cases, the partitive article is used without the definite article, e.g., if an adjective precedes the noun or in negative contexts: *de belles fleurs* 'beautiful flowers', *Elle n'a pas d'argent* 'She has no money / She doesn't have any money'. For an overview of the forms of the French articles, see table A2 in the appendix.

(12) a. (…) *et* **∅** *l'* *argent* (08 PT)
and the money
'… and (some) money'
[TL: *et* ***de l'****argent*]

b. *donc je ne* *mange pas* LIKE **∅** *la* *viande* (08 PT)
so I NEG eat not like the meat
'so, I don't eat meat'
[TL: *donc je ne mange pas* ***de*** *viande*]

The structure in (12a) has been produced in a context where the learner is listing presents they expect to receive for Christmas. Here, the learner inserts the early system morpheme definite article into a gap projected for the late system morpheme partitive article in an affirmative context referring to a part of a whole. In (12b), the learner uses the early system morpheme definite article in a negative context with the uncountable noun *viande* 'meat'. Both examples can be interpreted as instances of incomplete acquisition leading to the insertion of TL early system morphemes into late system morpheme gaps. However, it may also be argued that the examples in (12) represent cases of open gaps due to fossilization if it were the case that the learners have not yet learned that French requires the partitive article in these cases.

Example (13) shows that also early system morpheme verb inflections may be inserted into gaps for late system morpheme gaps. In (13a), the early system morpheme infinitive suffix *-er* is inserted into the gap projected for the third-person singular suffix *-e*. In (13b), the learner substitutes the subject-verb agreement marker *-t* with an early system morpheme past participle suffix which is underlined by their comment in their L1 English following the utterance. As the TL is not constrained in how it may fill in gaps in IL, both examples conform to the predictions of the CML model.

(13) a. *il* *retourn-****er*** *à* *la* *maison lentement* (08 NA)
he return-INF to the house slowly
'he returns to the house slowly'
[TL: *il retourn****e*** *à la maison lentement*]

b. *elle peind-****u*** *ou non*? I DON'T KNOW HOW TO SAY THAT
she paint-PPTCP or no?
IN PRESENT TENSE (08 LT)
'she is painting, isn't she?'
[TL: *elle pein****t****, n'est-ce pas*?]

However, it has to be noted here that the participle suffix *-u* is not the correct past participle suffix for the verb *peindre* whose irregular past participle form is *peint*, which is orthographically and phonologically identical to the third-person singular form of the verb in present

tense. Thus, the insertion of the irregular past participle suffix *-u* could not only represent TL overgeneralization but also an effect of instruction: French classes often put a special focus on the instruction of the different irregular past participle forms which could result in a higher level of activation of these forms in the language production process.

5.2.2.3 Target language late system morphemes in interlanguage gaps Apart from content morphemes and early system morphemes, late system morphemes from TL have also been overextended and inserted into gaps for early system morphemes in IL. For instance, this has happened with subject-verb agreement, as shown in (14), where the verb *être* is not correctly inflected for number. Here, it can be argued that the third-person singular form, which is already acquired by the learner, substitutes the appropriate third-person plural form which has not been acquired yet.

(14) *les touristes* ***est*** *très surpris* (19 LT)
the tourists be.3SG very surprised
'the tourists are very surprised'
[TL: *les touristes* ***sont*** *très surpris*]

Along with verb inflections for subject-verb agreement, late system morpheme subject clitics are subject to TL overgeneralization, as in (15). Unlike English, French has two series of personal pronouns, namely strong and clitic pronouns.[18] The strong pronouns occur in argument positions and carry thematic roles. The clitic pronouns, however, are semantically and syntactically deficient (Gabriel, Müller, & Fischer, 2018, pp. 155–157). At the semantic level, their referential scope is less restricted than that of personal pronouns: They may not only represent human referents but also inanimate arguments or complete predicates. At the syntactic level, clitics do not occur in argument positions or receive thematic roles. Rather, they are co-indexed with them and appear under INFL, acting like a part of agreement morphology (Jake, 1994, p. 273). Hence, French strong pronouns are classified as content morphemes under the 4-M model and clitic pronouns are categorized as outsider late system morphemes (Jake, 1994; Myers-Scotton & Jake, 2000).

(15) a. *qu'* *est-ce que* ***tu*** *as* *fait?* (25 QT)
what Q you have done
'what has he done?'
[TL: *qu'est-ce qu'****il*** *a fait*?]

une famille (...) ***il*** *arrive* *en une maison* (19 LT)
a.F family he arrives in a house
'a family arrives in a house'
[TL: *une famille (...)* ***elle*** *arrive dans une maison*]

[18] For an overview of the strong and clitic pronoun paradigms for subject, object, and reflexive clitics, see table A3 in the appendix.

The examples in (15) illustrate that in some cases, the wrong form of the subject clitic in terms of person or gender is selected. In (15a), the task for the learners was to ask about activities carried out by persons in pictures shown to them, requiring the use of the third-person singular clitic *il*. Instead, the gap for this late system morpheme is filled with the second-person singular clitic *tu*. In (15b), the default masculine form of the subject clitic *il* is used to refer to the subject DP *une famille* even though it should be feminine, in agreement with the gender of the noun. In both cases, it can be argued that the subject clitic paradigm has not been completely acquired by the learners which leads to the use of other forms from the same paradigm.

In addition, TL insertion has occurred with the bridge late system morpheme *de* in both its function as a partitive article and as a genitive marker, as demonstrated in (16). Similarly to the partitive article, the preposition *de*, marking possession or genitive, combines with the definite article to form different forms, i.e., *du*, *de la, de l'*, *des* meaning 'of the'. On a morphosyntactic level, it joins together two units (two NPs = the possessor and the possessed) and is therefore categorized as a bridge late system morpheme (Jones, 2018, p. 421).

(16) a. *le monstre de le lac Ness* (19 LT)
the monster of the lake Ness
'the monster of Loch Ness'
[TL: *le monstre* ***du*** *lac Ness*]

b. *le homme vole de tabac* (20 MT)
the man steals partART tobacco
'the man steals (some) tobacco'
[TL: *l'homme vole* ***du*** *tabac*]

In (16a), the possessive marker has not been contracted with the definite article *le* to form the contracted morpheme *du*. Instead, the bridge late system morpheme *de* and the early system morpheme definite article *le* have been inserted as separate morphemes. In (16b), the partitive article *de* is used without the definite article. Both cases represent instances of insertion of TL forms in a non-target-like way due to incomplete acquisition of the respective paradigms.

Furthermore, example (17) illustrates that late system morpheme prepositions from TL can be inserted into IL structures when the appropriate choice is not available to the learner yet. In French, the prepositions *à* and *de* are used to connect infinitives with the matrix CPs hosting them which qualifies these prepositions as bridge late system morphemes (Myers-Scotton & Jake, 2009, p. 349). In (17), the bridge late system morpheme preposition *de* required by the verb *décider* has been substituted by the preposition *à* which may also introduce infinite clauses in French but represents an inappropriate choice here. Thus, incomplete acquisition leads to the insertion of a TL bridge late morpheme in a non-target-like way.

(17) *il y a une famille qui* (…) *décidait* ***à*** *passer les vacances sur une lac* (20 LT)
it there has a family that decided to spend the holidays by a lake
'there is a family that decided to spend the holidays by a lake'
[TL: *il y a une famille qui a décidé* ***de*** *passer les vacances au bord d'un lac*]

Finally, a few examples seem to show insertion of TL late system morphemes into gaps for conceptually activated morphemes in IL, as in (18) and (19).

(18) *et* ***il***? (27 PT)
and he.CLIT?
'and (what about) him?'
[TL: *et* ***lui***?]

(19) *pour montr-****e*** *la mère et les autres enfants que le monstre du Loch Ness* (…) *existe vraiment* (21 LT)
to show-3SG the mother and the other children that the monster of-the Loch Ness exists really
'…to show the mother and the other children that the monster of Loch Ness really exists'
[TL: *pour montr****er*** *à la mère et aux autres enfants que le monstre du Loch Ness existe vraiment*]

In example (18), a late system morpheme, namely the third-person singular masculine subject clitic *il*, has been inserted into a gap where the TL would require a content morpheme, i.e., the strong personal pronoun *lui*, to be used. The L1 cannot serve as an explanation here because it does not differentiate between strong and weak pronouns and thus cannot influence the choice of the late system morpheme over the content morpheme in IL. As the CML model does not constrain the TL in the ways it may fill in gaps in IL, (18) does not represent a counter-example to the predictions made in the model.

In example (19), a late system morpheme subject-verb agreement inflection has been inserted into a gap for an early system morpheme infinitive suffix: The learner has produced the third-person singular *-e* instead of the infinitival *-er*. However, this also represents an instance of TL material filling in gaps in the learners' IL and thus does not contradict the predictions of the CML model.

5.2.2.4 Overgeneralization of abstract lexical information from the target language Some IL structures include superfluous morphemes from TL even though the gaps for these are not projected by abstract information underlying the lexical items from TL used in the utterances and neither influenced by L1 structure, as demonstrated in example (20). In (20a), a gap for an early system morpheme definite article accompanying the noun *voiture* is projected and filled with an early system morpheme from TL even though the TL does not require the overt presence

of the determiner in this construction. The L1 cannot be the source of the appearance of the article because the equivalent English structure does not make use of an article either. Similarly, in (20b), the direct object of the verb *finir* is introduced by an early system morpheme preposition even though neither TL nor L1 require a preposition in the realization of the object.

(20) a. *ils partir en* ***le*** *voiture* (19 LT)
they leave in the car
'they leave by car'
[TL: *ils partent en* ***Ø*** *voiture*]

b. *quand le grand-mère a fini* ***de*** *sa peinture* (24 LT)
when the grandmother has finished to her painting
'when the grandmother has finished her painting'
[TL: *quand la grand-mère a fini* ***Ø*** *sa peinture*]

The examples in (20) display TL overgeneralization not in the form of insertion of actual surface morphemes from TL but at a more abstract level. In both examples, it can be argued that abstract lexical structure from other TL morphemes or constructions is overextended. In (20a), the learner might have transferred the fact that French mostly requires the *phi*-features of nouns to be spelled out by the election of an article, which is not the case in the construction they intended to use. In (20b), the learner might not yet have acquired the information that the verb *finir* may either be complemented by an infinitival clause introduced by *de* or by a direct object DP without preposition, leading to the overextension of the prepositional construction used with infinitival complements to the direct object complement.

Up to this point, the IL structures discussed in section 5.2.2 illustrate that the insertion of TL material into IL gaps satisfies the complete-projection principle stated under the CML model. It also becomes clear that overgeneralization of TL forms and abstract structure does not necessarily lead to target-like IL structures but rather results in variation in IL. The next section addresses the occurrence of open gaps in IL which are also allowed as a case of fossilization under the CML model.

5.2.3 Open gaps in interlanguage

As predicted by Jake (1998, p. 363), instead of transferring abstract lexical structure from L1 or inserting TL surface morphemes, gaps in IL may also remain unfilled, assumably reflecting fossilization. Examples are shown in (21) for conceptually activated morphemes and in (22) for late system morphemes.

(21) a. *il s' appelle* (…) ***Ø*** *son petit ami*? (08 PT)
he himself is.called his boyfriend
'what is he called, her boyfriend?'
[TL: *il s'appelle* ***comment*** *son petit ami*?]

b. *j' ai visité ∅ Normandie* (23 PT)
I have visited Normandy
'I visited Normandy'
[TL: *j'ai visité **la** Normandie*]

(22) *il ∅ assis*? (21 QT)
he sat
'is he sitting?'
[TL: *il **est** assis*?]

In (21a), the gap for the content morpheme question word *comment* 'how' is left open. In (21b), the French early system morpheme definite article is missing. In French, the geographical name *Normandie* as a content morpheme requires the *phi*-features of gender and number to be spelled out explicitly and thus indirectly elects a direct article. In the example, the learner does not have this information available. As a result, the definite article is omitted in IL. Note that in the case of (21b), L1 influence cannot be ruled out because in English, the definite article would not be projected either, so that this case might also be attributable to L1 influence. Whether transfer or fossilization is the actual source of the structure cannot be determined because we cannot retrace the production of the structure in the learner's mind. In (22), the gap for the late system morpheme copula that is required for the grammaticality of the question is unfilled.

What the IL structures in (21) and (22) show is that in some cases, incomplete acquisition or unavailability of TL abstract structure or surface morphemes may result in the occurrence of open gaps which are neither filled with L1 structure nor TL elements and thus reflect fossilization. These gaps may eventually disappear when new TL structure is acquired, resulting in the projection of new, more target-like IL surface structures (Jake, 1998, p. 363).

In summary, the IL structures presented in chapter 5.2 illustrate how IL structures can be explained through L1 influence or TL overextension. Up until this point, the analysis has concentrated on examples confirming the prediction that IL surface structures can be explained as projections of composite lexical structure from L1 and TL (RQ2). Also, the examples discussed up to this point conform to the constraints on L1 influence (RQ 3). However, the discussion has remained selective with regard to the IL structures used as evidence until now, which does not yet prove that the principles underlying the CML model can account for all IL surface structures in the learner data.

5.3 Ambiguous cases and potential counter-examples

This chapter focuses on IL structures that cannot be explained by the CML model and the principles underlying it as straightforwardly as those discussed above (RQ 4). Firstly, section 5.3.1 discusses problems regarding the classification of morphemes under the 4-M model that have come up during the analysis, showing that the application of the criteria for the categorization of morphemes, which is necessary for working with the CML model, is far from trivial.

Secondly, section 5.3.2 addresses IL structures that are ambiguous with regard to what the learners intended to produce, shedding light on the challenge of working with spoken data. Thirdly, in section 5.3.3, the discussion turns to structures that cannot be explained through either L1 or TL insertion, pointing out that there are IL structures that the CML model cannot account for. Finally, section 5.3.4 presents IL structures which can be interpreted as counter-examples to the system-morpheme principle because they seem to demonstrate influence of L1 abstract lexical structure on TL late system morphemes.

5.3.1 Unclear status of morpheme type under the 4-M model

Some of the IL structures found in the data have not been discussed above because their status under the 4-M model is controversial. The example in (23), for instance, represents an utterance of a participant using an inappropriate preposition introducing the GOAL of a movement. The example is problematic because the status of the prepositions involved under the 4-M model is debatable.

(23) *peut-être* (...) *moi et ma famille* (...) *vont aller* (...) ***au*** *France* (21 PT)
maybe me and my family will go to.the France
'maybe me and my family will go to France'
[TL: *peut-être moi et ma famille (nous) irons* ***en*** *France*]

The French verb *aller* 'to go' is intransitive and can be accompanied by a locative adjunct PP introducing the GOAL of the movement. If the destination of the movement is a country, French has two different prepositions expressing 'to': *à* and *en* assigning the thematic role of GOAL to the noun contained in the respective locative PP. So far, the prepositions may be categorized as content morphemes (Myers-Scotton & Jake, 2009, p. 349).[19]

However, the distribution of *à* and *en* depends on the grammatical gender of the country referred to and sometimes on its phonological properties as well. The preposition *à* is the default choice used with masculine country names. The preposition *en* is used with country names that have feminine gender and/or start with a vowel or a silent *h*, e.g., *en Italie* 'to Italy', *en Hongrie* 'to Hungary'. Hence, the form of the preposition governing the adjunct PP only becomes salient when it is joined with the DP to build the complete PP. Therefore, the prepositions *à* and *en* assigning directionality to adjunct PPs could also be classified as bridge late system morphemes satisfying well-formedness requirements within their own maximal projection.[20] Still, as the example represents a case of insertion of TL elements into gaps in IL which is not constrained under the CML model, it does not represent a potential counter-example to the predictions made in the model.

[19]Note that the preposition *à* changes its form when it contracts with the masculine and plural form of the definite article: *à + le = au,* e.g., *au Mexique* 'to Mexico' and *à + les = aux*, e.g., *aux Pays-Bas* 'to the Netherlands'. As mentioned above, the discussion here follows Myers-Scotton and Jake (2009, p. 349) assuming that this process does not change the status of the morpheme from conceptually activated to structurally assigned.

[20]What is particularly puzzling here is the fact that the election of *en* may solely depend on phonological information concerning the noun which has not been addressed as relevant for the study of language contact on a morphosyntactic level by Myers-Scotton and Jake so far.

Another issue concerning the classification under the 4-M model is the status of the phrase *est-ce que* used to form questions in French, as in (24).

(24) a. *qu'* ***est-ce qu'*** *il a* *passé*? (21 PT)
what Q.OBJ it has happened
'what happened?'
[TL: *qu'est-ce* ***qui*** *s'est passé*?]

b. *qu'* ***est-ce qui*** *il* *apporte* (25 QT)
what Q.SBJ he wears
'what is he wearing?'
[TL: *qu'est-ce* ***qu'****il porte*?]

The French language has several options for forming questions that are related to different registers for both *yes/no*-questions and *wh*-questions. For instance, in formal speech, questions may be formed via subject-verb inversion while, in colloquial and informal speech, direct word order accompanied by a rising intonation pattern marking the question is a common strategy.

The third option, occurring in standard French, is the insertion of the phrase *est-ce que*. For *yes/no*-questions, the phrase occurs in a sentence-initial position, e.g., *Est-ce que tu me vois?* 'Do you see me?'. With *wh*-questions asking for the subject or the object of a sentence, the picture is more complex. Firstly, one of two question words, i.e., *qui* 'who' or *que* 'what', appears in the specifier position of CP. Secondly, the form of the phrase *est-ce que* alternates according to the syntactic status of the element that is asked for: *est-ce qui* is inserted if the question asks for the subject of a proposition; *est-ce que* is used when asking for an object.[21]

In (24), *est-ce qui*/*est-ce que* is used inappropriately. In (24a), the learner asks for a subject but uses *est-ce que* instead of *est-ce qui*. In (24b), the learner asks for the object using *est-ce qui* instead of *est-ce que*. The choice of the question word *que*, however, is correct in both examples.

While the status of the question words as content morphemes is relatively clear as they occur in argument positions, the classification of the phrase *est-ce qui*/*est-ce que* is more problematic. This is mainly due to controversy concerning the linguistic status of *est-ce que* itself. According to Rowlett (2007, p. 209), *est-ce qui*/*est-ce que* can be seen as grammaticalized complementizers in contemporary French that appear under C (the complementizer position).[22] He claims that the complementizers are inserted directly from the lexicon into their position under C, which could be used as an argument for their classification as content morphemes. However, they do not carry semantic information or thematic roles and their occurrence depends on information

[21] An overview of the formation of *wh*-questions with *est-ce que* in French is given in table A4 in the appendix.

[22] It could also be argued that the construction of *wh*-questions using *qui*/*que* and the question marker *est-ce qui*/*est-ce que* is an instance of pronominal inversion, e.g., *C'est qui que tu vois*? 'It is who that you see?' > *Qui est-ce que tu vois*? 'Who is-it that you see?'. An exhaustive discussion of these two interpretations is beyond the scope of this thesis. As in standard French, the phrase *est-ce qui*/*est-ce que* is insensitive to tense or person/number agreement, it is treated as a fixed complementizer stored in the lexicon here. For a detailed discussion of both views, see Rowlett (2007, pp. 209–214).

that only becomes available when the clause itself is assembled, as they mark the interrogative nature of a sentence. Thus, they are accessed at the level of the Formulator and qualify as late system morphemes.

What remains puzzling is the mechanism that leads to the choice of either *est-ce qui* or *est-ce que*: If they are retrieved as ready-made chunks and directly inserted under C, how can they be sensitive to the syntactic status of the element asked for in the question and change their form accordingly? This question could be a starting point for further research into the application of the 4-M model concerning question formation in French.

Regardless of the difficulties concerning the classification of *est-ce qui*/*est-ce que* under the 4-M model, the examples in (24) do not represent potential counter-examples to the CML model. The L1 English does not dispose of an equivalent phrase for marking the syntactic status of the questioned element and is thus not an option as a possible source for the inappropriate choices in (24). As a result, it can be concluded that the IL structures in (24) display instances of insertion of TL material due to incomplete acquisition which is not constrained under the 4-M model.

Notwithstanding their compatibility with the predictions made by the CML model, the examples discussed in this section demonstrate that a very detailed analysis and constant consideration of language-specific features are necessary to apply the 4-M model to different languages. Also, it has become clear that further work has to be done here to make the principles of the MLF model and its sub-models applicable to a broad range of language-contact data with different participating languages.

5.3.2 Ambiguous structures

Ambiguity in the examples has been caused by phonological similarities between verb inflections in French. Particularly, the verb inflections marking the infinitive (*-er*) and regular past participle (*-é*) are pronounced identically as [e] and similarly or even identically to the imperfect inflections for first- to third-person singular and third-person plural (*-ais, -ais, -ait, -aient*) which are pronounced as the more open [ɛ], even though native speakers may also tend to produce [e] instead. Therefore, in utterances like those cited in (25), it becomes almost impossible to determine with certainty which tense a speaker intended to produce.

In (25), the learner describes an ongoing action in the past. In French, this would require the use of the imperfect form of the verb, i.e., *dessinait* [desinɛ]. One possible explanation of the structure in (25) is that the learner has correctly chosen the imperfect tense which they realized as [e] instead of [ɛ]. However, they could also have resorted to the infinitive form (*dessiner* [desine]) of the verb, representing the insertion of an early system morpheme verb into the late system morpheme gap for the imperfect inflection. The hesitation preceding the actual production of the verb emphasizes the learner's uncertainty regarding the correct form or pronunciation of the correct inflection for the verb. Despite its ambiguity, this example is not problematic with regard to the predictions made under the CML model, as the TL is not constrained in filling in gaps in IL.

(25) *le* fils (…) *dessin-**er***[23] *une image du lac* (24 LT)
the son draw-? a picture of.the lake
'the son was drawing an image of the lake'
[TL: *le fils **dessinait** une image du lac*]

5.3.3 Structures not explicable through either the native or the target language

The main assumption of the CML model is that IL structures can be explained as projections of composite lexical structure from L1 and TL. However, the data include structures for which this prediction is not true. The structures in the data that are not explicable as based on abstract lexical structure from either L1 or TL all concern word order.

In (26), the reflexive clitic *se* follows the finite verb and precedes the participle, even though in French, the clitic occurs under INFL and therefore precedes the finite verb form. The L1 cannot be the source here because, in English, the reflexive pronoun follows both the finite verb and the participle in composite tenses, e.g., *They saved themselves*. However, the structure in (26) represents a single case and might also represent a performance error. Here, a larger amount of data is required to find out whether this kind of word order variation occurs systematically in the IL of English learners of French.

(26) *ils **sont se** sauvé* (28 MT)
they are themselves saved
'they saved themselves'
[TL: *ils **se sont** sauvés*]

Another construction involving word order variation that cannot be traced back to abstract lexical structure from either L1 or TL is the negation. Negation in French involves the negator *ne* preceding the finite verb and the negative adverb *pas* 'not' following the finite verb. In a negated sentence, subject clitics precede the negation phrase, object and reflexive clitics directly follow the negator *ne* and precede the finite verb followed by *pas*, and full object DPs and infinitive verbs or participles follow *pas*. While the negator *ne* may be omitted in surface structures in colloquial French, the negative adverb *pas* is always required to be overtly realized.

The IL structures involving negation in (27) exhibit a considerable amount of variation concerning word order. In (27a), *pas* precedes the finite verb instead of following it. In (27b), the object DP *l'ordinateur* precedes *pas* instead of following it. In (27c), the negator *ne* follows the reflexive clitic *se* instead of preceding it.

[23] Throughout the transcriptions, the production of non-target-like [e] by learners has been transcribed as *-er*. In past-tense contexts, the transcribers have attempted to listen to the difference between [e] and [ɛ] (Myles & Mitchell, 2013f).

(27) a. *elle ne* ***pas brosse*** *les dents* (30 NA)
she NEG not brushes the teeth
'she is not brushing her teeth'
[TL: *elle ne se* ***brosse pas*** *les dents*]

b. *elle n'* *utilise* ***l'*** ***ordinateur*** *pas* (21 NA)
she NEG uses the computer not
'she is not using the computer'
[TL: *elle n'utilise* ***pas l'ordinateur***]

c. *elle* ***se*** ***ne*** *lève* *pas vite* (28 NA)
she herself NEG get.up not quickly
'She is not getting up quickly'
[TL: *Elle* ***ne se*** *lève pas vite*]

In addition to these cases of word order variation, the negative adverb *pas* is omitted very frequently in the data. The L1 cannot be the source of the variation concerning word order and the omission of *pas* because the English negation involves only one negative element (*not*) which directly follows the finite verb and precedes objects and infinitives. The TL cannot be the source either because there are no negative constructions in standard French where the finite verb may follow *pas,* the object DP may precede *pas,* or the reflexive clitic may precede the negator *ne.* Therefore, these examples illustrate that there are structures in learners' interlanguage that cannot be accounted for under the CML model. At this point, we can only make educated guesses as to how these structures may have arisen: They could, for example, be the result of compromise strategies such as trial and error, or just performance errors related to limited working memory capacities that make it difficult to deal with negative clauses involving several arguments and clitics in addition to the two elements of the negation itself.

In the future, it would be interesting to conduct studies including larger databases to investigate whether these patterns of word order variation in negation are systematic competence errors or performance errors and whether potential systematic patterns may be explained by either L1 or TL influence. Also, other L1s than English could be included to see whether different L1 negation patterns, e.g., pre-verbal, post-verbal, or double negation, may influence IL word order patterns in negation.

5.3.4 Native language structure affecting late system morphemes in interlanguage

In addition to IL structures that cannot be accounted for by either L1 transfer or insertion of TL elements, some examples seem to display violations of the constraints on the possible contribution of the L1 in constructing IL. According to the system-morpheme principle, the L1 may only provide abstract lexical structure underlying content morphemes and early system morphemes in the L1 which may only fill in gaps projected for content morphemes and early system morphemes in IL. The examples in (28) show that abstract lexical structure underlying

content morphemes in the L1 may influence the realization of late system morpheme clitic pronouns in IL.

(28) a. *il* ***suit*** ***le*** (08 MT)
he follows him
'he follows them'
[TL: *il* ***les suit***]

b. *elle n'* ***Ø*** *habille pas très vite* (08 NA)
she NEG dresses not very fast
'she is not dressing (herself) very fast'
[TL: *elle ne* ***s'****habille pas très vite*]

In (28a), the word order is concerned: The direct object clitic *les* 'them' (which is also erroneously used in its singular form *le* even though it refers to two persons) follows the finite verb instead of preceding it. This can be attributed to the L1 English where the object pronoun *them* also follows the finite verb, that is, the morphological-realization patterns underlying an object pronoun which is a content morpheme in the L1 have been transferred to an outsider late system morpheme clitic in TL, thus violating the system-morpheme principle.

The example in (28b) is related to the overt realization of reflexive clitics. In French, verbs expressing actions of the monadic personation type, i.e., where the actor carrying out the action is also the patient of the action, require the overt realization of a reflexive clitic. In English, in contrast, monadic personation may be expressed by a simple verb where the reflexive pronoun can be overtly expressed for emphasis but is not necessary: e.g., *se raser* 'to shave (oneself)' (Talmy, 2000, pp. 89–90).

Under the 4-M model, French clitics are outsider late system morphemes. This is also valid for reflexive clitics. They occur under INFL, manipulating the predicate-argument structure of the verb they occur with, i.e., reducing the number of arguments by one and marking coreferentiality of PATIENT and AGENT (Jackendoff, 1990, p. 68). French reflexive clitics are coindexed with an element outside their own maximal projection, namely the subject. In (28b), the morphological-realization patterns underlying reflexive pronouns in the L1 English, i.e., the information that they do not necessarily have to be overtly realized, is transferred to the morphological-realization patterns underlying the realization of the late system morpheme reflexive clitics in IL, which represents a violation of the system-morpheme principle.

However, the structures in (28) might be explicable by considering reanalysis as a possible mechanism interacting with SLA and transfer. As Jake (1998, p. 342) puts it, "real or putative" lexical structure from TL must underlie all IL structures. In (28), this means that we cannot rule out the possibility that the learners have reanalyzed the French late system morpheme clitics as content morpheme strong pronouns. This possibility is also supported by Gross (2000) who argues that morphemes, which have a multiple status in a language, i.e., which may be classified as conceptually activated or structurally assigned depending on the context they occur in, may be reanalyzed as conceptually activated in all contexts by individual learners.

In this case, the structures in (28) would not violate the system-morpheme principle as the TL morphemes that the abstract information from L1 content morphemes is inserted into would be classified as content morphemes by the learners' individual grammars.

Furthermore, (29) demonstrates that the realization of outsider late system morpheme auxiliary verbs may be affected by L1 abstract lexical structure. Here, the *passé composé* form of the motion verb *arriver* is constructed with an inflected form of the auxiliary *avoir* 'have' like in the L1 English although French requires the *passé composé* of motion verbs to be constructed using *être* 'be'.

(29) *beaucoup de* *personnes* ***a*** *arrivé* (21 LT)
many partART persons has arrived
'many persons arrived'
[TL: *beaucoup de personnes* ***sont*** *arrivées*]

The auxiliary that is used in the construction of the *passé composé* is classified as an outsider late system morpheme with reference to Myers-Scotton (2008, pp. 34–35) who states that morphemes expressing tense are outsider late system morphemes when they interact with agreement. As the auxiliary in the *passé composé* construction inseparably expresses tense and realizes subject-verb agreement, the complete form is categorized as a late system morpheme. Also, Wei (2000a, p. 33) notes that auxiliaries are late system morphemes when they are required by the grammar, which is the case with auxiliaries used to construct the *passé composé* form of the verb *arriver*. Thus, in (29), it seems that abstract lexical structure from the L1 English affects the form of a late system morpheme in IL. This, in turn, would violate the system-morpheme principle.

However, it could also be argued that the auxiliaries *avoir* and *être* used to form the *passé composé* are multimorphemic lexemes. According to Myers-Scotton (2008, p. 34), tense and aspect are early system morphemes when they do not interact with agreement. As a consequence, from the perspective of language production, *avoir* and *être* might start as early system morphemes that are indirectly elected by the lexical verb to express aspect and whose final form does not become salient until the level of the Formulator when subject-verb agreement is realized. While this does not change their ultimate status as outsider late system morphemes from a classificational perspective, it sheds light on the possibility that what is actually influenced by the L1 in (29) is information underlying the election of the early system morpheme auxiliary verb and not information affecting subject-verb agreement. Under this perspective, the example in (29) would conform to the predictions made under the CML model.

Finally, in (30), the occurrence of the late system morpheme copula *être* seems to be based on L1 structure. In both examples, the learners may have aimed to express progressive aspect. In (30a), an inflected form of the auxiliary *être* is combined with the infinite verb *courir* to convey the ongoing action of running. In (30b), the inflected auxiliary *être* is combined with the verb *danser*, which is also inflected for person and number, expressing the ongoing action of dancing.

(30) a. *les hommes* ***sont cour-ir*** (23 NA)
the men are run-INF
'the men are running'

[TL: *les hommes* ***courent*** / *les hommes* ***sont en train de courir***]

b. *il* ***est dans-e*** *à la musique* (08 NA)
he is dance-3SG to the music
'he is dancing to the music'
[TL: *il* ***danse*** *sur la musique* / *il* ***est en train de danser*** *sur la musique*]

The structures in (30) could be the results of attempts to transfer the English progressive construction into IL. In French, progressive aspect is usually not overtly marked. Ongoing actions are either expressed in present tense, e.g., *Il court* 'He runs / He is running' or by resorting to the somewhat cumbersome expression *être en train de* + INF, e.g., *Il est en train de lire* 'He is in the process of reading'. The learners in (30) do not appear to have full access to these possibilities of expressing ongoing actions in French. Hence, they may have resorted to their L1 English where progressive aspect in the present tense is expressed by combining the late system morpheme auxiliary *be*, which has to be co-indexed with the subject in person and number, with the present participle form of the lexical verb, e.g., *He is running*. As a result, abstract lexical structure from the L1 may have acted as the basis for projecting the gaps for the outsider late system morpheme auxiliary *être* in (30). These gaps may then have been filled with the respective TL surface forms. It could thus be argued here that the L1 affects late system morphemes in IL, which would represent a further violation of the system-morpheme principle in the data analyzed in this study.

However, it is not necessarily the case here that the auxiliary late system morphemes have been directly projected by the L1. Rather, it could be argued that the concept of progressive itself, including the information that progressive aspect has to be spelled out in present tense, has been transferred from the L1 to IL while the gaps projected for the late system morpheme verb inflections for person and number are still provided by TL (cf. Jake, 1998, p. 366 for a similar argument regarding the overuse of aspect marking by Arabic learners of English). Note that the transfer is not complete as the lexical verbs do not display participle inflections as it would be the case in English, which strengthens the argument that the gaps for the verb inflections involved in marking progressive aspect in English have not directly been transferred to the learners' respective IL frames. Thus, the structures in (30) do not necessarily represent counter-examples to the system-morpheme principle even though they involve late system morphemes influenced by abstract lexical structure from the L1.

In sum, the IL data discussed in chapter 5.3 illustrate that there are instances of IL structures posing problems concerning the status of the morphemes involved under the 4-M model and the learners' intentions, as well as IL structures that cannot be accounted for under the 4-M model. In some cases, IL late system morphemes, namely auxiliaries and pronominal cli-

tics, seem to be affected by abstract structure from the L1 which would violate the system-morpheme principle and thus might represent counter-examples to the CML model (even though alternative explanations are possible).

However, it should be noted that these potential counter-examples are small in number, whereas the large majority of IL structures found in the data are in line with the predictions made under the CML model. As Mindt (2002, pp. 199–210) demonstrates, grammatical rules mostly display a low amount of exceptions of up to 5% of all cases when tested against language data. These exceptions may involve errors or slips of the tongue as well as idiosyncrasies and obsolete or innovative structures (Mindt, 2002, p. 211).

Therefore, the mere existence of these problematic structures should not be taken as counterevidence to the explanatory power of the model in general. Rather, these examples show that learner grammars are dynamic and flexible and that production does not always reflect the abstract grammar of individual learners. As we deal with performance data here, it could also always be the case that learners produce the same morphemes and constructions correctly in other situations, or that similar deviations might occur in the speech of native speakers in comparably small numbers.

Still, the existence of potential counter-examples to the CML model should not be completely neglected. To determine whether they are significant in number and therefore impact the predictive power of the model, further research involving larger collections of data from different languages, targeted elicitation of relevant constructions, and quantitative analyses are needed.

6 Conclusion and outlook

This thesis investigated whether the predictions made under the CML model of SLA (Jake, 1998) can be confirmed with data from English learners of French as an L2 as a test of the claim that the model generalizes to SLA data independently of the language pair involved. After defining the central concepts of SLA, transfer, and interlanguage (chapter 2), an introduction into the theoretical underpinnings of the CML model was given (chapter 3.1) and the resulting predictions for interlanguage development were outlined (chapter 3.2). An overview of previous research indicated that the principles of language contact stated under the MLF model generalize to different language-contact situations, justifying its application to SLA data. It also demonstrated that while there is data providing evidence for the predictions of the CML model, the available evidence is selective and not representative (chapter 3.3). This led to the formulation of the concrete research question guiding the analysis (chapter 3.4). After the presentation of the interlanguage data and the method of analysis (chapter 4), the results of the analysis were discussed with regard to their implications for the predictions of the CML model (chapter 5).

The analysis of the learner data has demonstrated that the predictions made under the CML model can be confirmed to a large extent. In chapter 5.1, it was shown that except for cases of codeswitching and loan blends, all IL structures are based on TL material at the surface and abstract level, confirming the target-language principle. Chapter 5.2 illustrated that IL

structures can be explained as projections of composite lexical structure from L1 and TL. Here, it was demonstrated that L1 influence takes place at all three levels of abstract lexical structure. It has also become clear that TL overgeneralization took place at the abstract as well as at the surface level and concerned content morphemes, early system morphemes, and late system morphemes, confirming that the insertion of TL material to fill in gaps is not constrained. These results confirm the complete-projection principle. Also, in the cases discussed in chapter 5.2, L1 influence was constrained to transfer from L1 content morphemes and early system morphemes into gaps projected by content morphemes and early system morphemes in IL as predicted by the system-morpheme principle.

However, there were IL structures that were problematic for different reasons. These have been discussed in chapter 5.3. Firstly, some non-target-like IL structures involved morphemes whose status under the 4-M model was unclear or structures that were ambiguous with regard to what the learner intended to produce (sections 5.3.1 and 5.3.2). These cases show that the application of the model is not always straightforward and that extensive and very detailed knowledge of the languages involved is required to properly and adequately apply it to learner data. They also demonstrate that sometimes, ambiguity cannot be resolved in the data due to phonological similarities between verb inflections in French and the fact that corpus data do not allow to draw conclusions on actual the speaker intentions underlying production processes. Secondly, IL structures involving word order variation represented potential counter-examples to the CML model in the way that they could not be explained as projections of composite lexical structure from L1 and TL because the varying order could not be attributed to either L1 or TL influence (section 5.3.3). However, it was noted that more data is needed to determine whether some of these deviations just represent performance errors or whether there are systematic patterns of word order variation which might be related to either L1 transfer, TL overextension, or other strategies such as the avoidance of transfer or trial and error.

Finally, there were non-target-like IL structures that seemed to directly violate the system-morpheme principle (section 5.3.4). These structures involved abstract lexical structure from L1 filling gaps projected for late system morphemes, such as clitic pronouns and auxiliaries. However, these instances of IL do not necessarily represent counter-examples to the system-morpheme principle, depending on the line of argumentation. On the one hand, it was argued that these IL structures can also be accounted for under the CML model by considering reanalysis and transfer at different levels of multimorphemic elements as alternative explanations. However, it has to be noted here that the suggested alternative explanations for the IL structures displaying possible L1 influence on late system morphemes in IL cannot be proven because no information about the actual individual learners' grammars is available. On the other hand, as noted in chapter 5.3, the instances of IL structures potentially violating the system-morpheme principle have been small in number. Following Mindt (2002), they might thus represent performance errors which might also be found in L1 speakers. In the future, quantitatively oriented research is required to determine whether these violations of the system-morpheme principle are significant and systematic.

As a summary, the following statements can be made with regard to the research questions

that guided the analysis:

1. The IL surface structures observed in the learner data are based on TL lexical items. The only exceptions are instances of codeswitching and loan blends which are permitted under the target-language principle. Hence, RQ 1 can be confirmed on the basis of the data.

2. Most of the IL surface structures observed in the data can be explained as projections of composite lexical structure from the L1 and TL. However, there are also cases of word order variation that cannot be explained through either L1 or TL influence. On the basis of the data analyzed in this study, it cannot be determined whether these cases represent competence errors or performance errors. Thus, RQ 2 can, to a large extent, be confirmed on the basis of the data, even though further research is required to find out whether potential systematic word order deviations may be explained by either L1 or TL influence.

3. The restrictions for the limited influence of the L1 hold for a large part of the IL surface structures observed in the learner data, but there exists a number of examples where abstract lexical structure from the L1 seems to affect late system morphemes in IL. However, these potential violations of the system-morpheme principle are small in number and may be explained by taking into consideration phenomena such as multimorphemic elements and reanalysis. Therefore, RQ 3 can be confirmed to a large extent, too, even though further research is needed to determine whether the potential counter-examples represent idiosyncratic structures or systematic patterns.

4. The examples of word order variation not attributable to either L1 or TL and the potential instances of L1 influence on late system morphemes in IL represent cases that are not explicable under the CML model straightforwardly. However, these cases are small in number and might also constitute performance errors. With regard to RQ 4, it can thus be stated that the analysis did not reveal any direct counter-examples to the predictions of the CML model, even though quantitative analyses are necessary to find out whether potential counter-examples exist as systematic patterns and in statistically significant numbers.

In the present study, it was not always clearly determinable whether a specific structure was based on L1 influence or insertion of TL material. This can be attributed to the fact that French and English are similar in many respects and at different levels of abstract lexical structure as well as at the surface level. Thus, future research into the CML model should include SLA data where the L1 and TL are typologically different to be able to separate L1 influence from TL overextension more clearly. Related to this point is the remark that the CML model does not predict which of the two strategies proposed to fill in gaps in IL, namely L1 transfer and TL insertion, is preferred by learners in situations where both are possible. With regard to these questions, factors such as the perceived distance between the languages involved, the level of proficiency, effects of instruction, individual learning strategies, and attitudes toward transfer

in general could be taken into account. Thus, the CML model could be related to other models and theories of SLA to find out whether they can complement each other in the explanation of IL structures.

In addition to these limitations regarding the analysis of the data, the CML model was criticized on a methodological level in this thesis. For instance, as already mentioned in section 3.1.3, the concept of sufficient congruence with acquired TL structure as a precondition for transfer of L1 structure or insertion of TL material is problematic. Apart from the lack of clarity in its definition, which is admitted by Myers-Scotton herself, this concept cannot be disproved empirically as we cannot infer from the data when exactly transfer has been blocked in the first place – whether by insufficient congruence or not.

Furthermore, the predictions made under the CML model heavily rely on the concept of individual learner grammars which differ from the actual TL norm to various extents. While this idea captures phenomena typical for SLA, such as reanalysis, non-native-like attainment, and fossilization, it makes it hard to falsify the predictions of the CML model. Specifically, the system-morpheme principle constraining L1 influence predicts that abstract lexical structure from the L1 may only be inserted into gaps that learners' individual grammars specify as projected by TL content morphemes. This notion may be invoked to explain structures that might otherwise represent counter-examples, as demonstrated in section 5.3.4. However, this line of argumentation results in circularity when the deviation of a learner's grammar from TL norms is firstly derived from data that does not conform to the system-morpheme principle and is then used to explain the same data.

In sum, while the CML model can account for a large amount of IL structures, it often relies on post-hoc interpretations. This weakens its explanatory power and its falsifiability. Thus, it would be desirable to test the predictions made under the CML model against research that independently taps into learners' individual IL grammars. In this regard, a promising starting point could be research projects such as the *English Profile Programme* (Hawkins & Buttery, 2010; Hawkins & Filipović, 2012; Milanovic, 2009) aiming to identify the grammatical properties of English interlanguage at different levels of acquisition in learners with different L1s. These independently established and statistically significant features of English interlanguage represent a solid database for examining the predictions of the CML model that is free from performance errors or individual patterns. In addition, comparing the features of English interlanguage in learners with different L1s would open up the possibility to determine with more certainty whether particular patterns in IL are due to L1 influence or TL overextension in cases where this would otherwise not be possible because of the similarities between the participating languages.

All in all, the main goal of the CML model of SLA by Jake (1998) is to describe, predict, and constrain which structures may arise in IL development. Summarizing the results of the analysis, it can be stated that the model can indeed account for most of the non-target-like IL structures found in the interlanguage data of English learners of French that were investigated here. Thus, this thesis adds to the evidence confirming the general applicability of the CML model to different L1-TL pairs. This result also further strengthens the relevance of the universal principles of language contact stated under the MLF model and their predictive

power for mono- and bilingual speech. However, it has become clear that the evidence for the universal applicability of the CML model of SLA is still far from representative and that larger-scale studies and quantitative analyses of interlanguage data involving typologically different L1-TL pairs are needed in the future to definitively confirm the predictions made under the model.

References

Abney, S. P. (1987). *The English noun phrase in its sentential aspect* (Dissertation). Massachusetts Institute of Technology (MIT).

American Psychological Association (2020). Gender. Retrieved from `https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/gender`

Bailey, N., Madden, C., & Krashen, S. (1974). Is there a 'natural sequence' in adult second language learning? *Language Learning*, *24*, 235–243.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459–484. `https://doi.org/10.1177/0267658307080557`

Batchelor, R. E., & Chebli-Saadi, M. (2011). *A reference grammar of French*. Cambridge: Cambridge University Press.

Bentahila, A. (1995). Review on: Duelling Languages. Grammatical structure in codeswitching. *Language*, *71*(1), 135–140. `https://doi.org/10.2307/415966`

Bolonyai, A. (1998). In-between languages: Language shift/maintenance in childhood bilingualism. *International Journal of Bilingualism*, *2*(1), 21–43. `https://doi.org/10.1177/136700699800200102`

Bolonyai, A. (2000). "Elective affinities": Language contact in the abstract lexicon and its structural consequences. *International Journal of Bilingualism*, *4*(1), 81–106. `https://doi.org/10.1177/13670069000040010601`

Bolonyai, A. (2002). Case systems in contact: Syntactic and lexical case in bilingual child language. *Southwest Journal of Linguistics*, *21*(2), 1–36.

Boussofara-Omar, N. (2003). Review on: Contact linguistics. Bilingual encounters and grammatical outcomes. *Bilingualism: Language and Cognition*, *6*(1), 33–46. `https://doi.org/10.1017/S1366728903001032`

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures* (7th ed.). Berlin: Mouton De Gruyter.

Corder, S. P. (1967). The significance of learner's [sic!] errors. *International Review of Applied Linguistics*, *5*, 161–169.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs—A case for contrastive analysis. *Studies in Second Language Acquisition*, *7*(1), 73–79. `https://doi.org/10.1017/S0272263100005167`

Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, *23*(2), 245–258.

Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*(1), 37–53.

Dulay, H., & Burt, M. (1975). Creative construction in second language learning and teaching. In H. Dulay & M. Burt (Eds.), *New directions in second language learning, teaching, and bilingual education* (pp. 21–32). Washington, D.C.: Teachers of English to Speakers of Other Languages.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two.* New York: Oxford University Press.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden: Blackwell.

Falk, Y., & Bardel, C. (2010). The study of the role of the background languages in third language acquisition. The state of the art. *International Review of Applied Linguistics in Language Teaching*, *48*(2-3), 185–219. https://doi.org/10.1515/iral.2010.009

Foley, C., & Flynn, S. (2018). The role of the native language. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 97–113). Cambridge: Cambridge University Press.

Fries, C. C. (1945). *Teaching and learning English as a foreign language.* Ann Arbor: University of Michigan Press.

Fuller, J. M. (1996). When cultural maintenance means linguistic convergence: Pennsylvania German evidence for the Matrix Language Turnover hypothesis. *Language in Society*, *25*(4), 493–514. https://doi.org/10.1017/S0047404500020790

Fuller, J. M. (1999). Between three languages: Composite structure in interlanguage. *Applied Linguistics*, *20*(4), 534–561. https://doi.org/10.1093/applin/20.4.534

Fuller, J. M. (2000). Morpheme types in a matrix language turnover: The introduction of system morphemes from English into Pennsylvania German. *International Journal of Bilingualism*, *4*(1), 45–58. https://doi.org/10.1177/13670069000040010401

Fuller, J. M., & Lehnert, H. (2000). Noun phrase structure in German-English codeswitching: Variation in gender assignment and article use. *International Journal of Bilingualism*, *4*(3), 399–420. https://doi.org/10.1177/13670069000040030601

Gabriel, C., Müller, N., & Fischer, S. (2018). *Grundlagen der generativen Syntax: Französisch, Italienisch, Spanisch* (3rd ed., updated and extended). Berlin, Boston: De Gruyter.

Gass, S. M. (2013). *Second language acquisition: An introductory course* (4th ed.). New York: Routledge.

Gass, S. M., & Selinker, L. (1993). Introduction. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 1–17). Amsterdam, Philadelphia: John Benjamins.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Progress*, *3*, 233–274.

Gross, S. (2000). When two become one: Creating a composite grammar in creole formation. *International Journal of Bilingualism*, *4*(1), 59–80. https://doi.org/10.1177/13670069000040010501

Hawkins, J. A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, *1*(1), 1–23. https://doi.org/10.1017/S2041536210000103

Hawkins, J. A., & Filipović, L. (2012). *Criterial features in L2 English: Specifying the reference levels of the Common European Framework. English profile studies: Vol. 1.* Cambridge: Cambridge University Press.

Jackendoff, R. (1990). *Semantic structures. Current studies in linguistics: Vol. 18.* Cambridge, MA, London: MIT Press.

Jake, J. L. (1994). Intrasentential code switching and pronouns: On the categorial status of functional elements. *Linguistics*, *32*, 271–298.

Jake, J. L. (1998). Constructing interlanguage: Building a composite matrix language. *Linguistics*, *36*(2), 333–382. https://doi.org/10.1515/ling.1998.36.2.333

Jones, M. C. (2018). Does language loss follow a principled structural path? Evidence from Jersey Norman French. *Journal of French Language Studies*, *28*(3), 399–429. https://doi.org/10.1017/S0959269518000042

Keller, M. (2020). *Code-switching: Unifying contemporary and historical perspectives. New approaches to English historical linguistics.* Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-34667-6

Klein, W., Dietrich, R., & Noyau, C. (1993). The acquisition of temporality. In C. Perdue (Ed.), *Adult language acquisition: Cross-linguistic perspectives* (2: The Results, pp. 73–118). Cambridge: Cambridge University Press.

Klein, W., & Perdue, C. (1993). Utterance structure. In C. Perdue (Ed.), *Adult language acquisition: Cross-linguistic perspectives* (pp. 3–39). Cambridge: Cambridge University Press.

Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning*, *27*(1), 93–107. https://doi.org/10.1111/j.1467-1770.1977.tb00294.x

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* London: Longman.

Lado, R. (1957). *Linguistics across cultures.* Ann Arbor: University of Michigan Press.

Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, *15*(1), 35–48. Retrieved from http://www.jstor.org/stable/44487587

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, Mass., London: MIT Press.

Liu, X. (2015). Language transfer in learning Japanese and interlanguage development. *US-China Foreign Language*, *13*(4), 234–244. https://doi.org/10.17265/1539-8080/2015.04.002

MacWhinney, B. (Ed.) (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah: Lawrence Erlbaum.

Meechan, M. (1995). Review on: Duelling languages: Grammatical structure in codeswitching. *Anthropological Linguistics*, *7*(1), 106–109.

Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, *3*, 109–135.

Milanovic, M. (2009). Cambridge ESOL and the CEFR. *Cambridge ESOL Research Notes*. (37), 2–5. Retrieved from http://www.cambridgeenglish.org/images/23156-research-notes-37.pdf

Mindt, D. (2002). What is a grammatical rule? In L. E. Breivik & A. Hasselgren (Eds.), *From the COLT's mouth ... And others': Language corpora studies in honour of Anna-Brita Stenström* (pp. 197–212). Amsterdam, New York: Rodopi.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. New York: Routledge.

Mpiranya, F. (2014). *Swahili grammar and workbook*. London: Routledge.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Muysken, P., & Rooij, V. de (1995). Review on: Social motivations for codeswitching. Evidence from Africa / Duelling languages. Grammatical structure in codeswitching. *Linguistics: An International Review*, *33*(5), 1043–1051.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

Myers-Scotton, C. (1997). *Duelling languages: Grammatical structure in codeswitching* (2nd ed.). Oxford: Clarendon Press.

Myers-Scotton, C. (1998). A way to dusty death: The Matrix Language turnover hypothesis. In L. A. Grenoble & L. J. Whaley (Eds.), *Endangered languages: Language loss and community response* (pp. 289–316). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166959.013

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes. Oxford linguistics*. Oxford: Oxford University Press.

Myers-Scotton, C. (2008). Language contact: Why outsider system morphemes resist transfer. *Journal of Language Contact, Thema 2*, 21–41.

Myers-Scotton, C. (2010). The matrix language frame model: Development and responses. In R. Jacobson (Ed.), *Codeswitching worldwide II* (pp. 23–58). Berlin, New York: Mouton De Gruyter.

Myers-Scotton, C., & Jake, J. L. (1995). Matching lemmas in a bilingual language competence and production model: Evidence from intrasentential code switching. *Linguistics*, *33*(5), 981–1024. https://doi.org/10.1515/ling.1995.33.5.981

Myers-Scotton, C., & Jake, J. L. (2000). Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics*, *38*, 1053–1100.

Myers-Scotton, C., & Jake, J. L. (2009). A universal model of code-switching and bilingual language processing and production. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 336–357). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511576331.020

Myers-Scotton, C., & Jake, J. L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, *21*(3), 340–366. https://doi.org/10.1177/1367006915626588

Myles, F., & Mitchell, R. (2013a). French learner language oral corpora (FLLOC). Retrieved from http://www.flloc.soton.ac.uk/

Myles, F., & Mitchell, R. (2013b). Newcastle Corpus: Database content. Retrieved from http://www.flloc.soton.ac.uk/newcastle/dbcontent.html

Myles, F., & Mitchell, R. (2013c). Newcastle Corpus: Description. Retrieved from http://www.flloc.soton.ac.uk/newcastle/index.html

Myles, F., & Mitchell, R. (2013d). Newcastle Corpus: Learners. Retrieved from http://www.flloc.soton.ac.uk/newcastle/learners.html

Myles, F., & Mitchell, R. (2013e). Newcastle Corpus: Tasks. Retrieved from http://www.flloc.soton.ac.uk/newcastle/tasks.html

Myles, F., & Mitchell, R. (2013f). Newcastle Corpus: Transcription conventions. Retrieved from http://www.flloc.soton.ac.uk/newcastle/conventions.html

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 436–486). Malden: Blackwell. https://doi.org/10.1002/9780470756492.ch15

Odlin, T., & Yu, L. (2016). Introduction. In L. Yu & T. Odlin (Eds.), *New perspectives on transfer in second language learning* (pp. 1–16). Bristol: Multilingual Matters.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, *6*, 186–214.

Pienemann, M. (1998). *Language processing and second language acquisition: Processability theory*. Amsterdam: John Benjamins.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish *y termino en español*: Toward a typology of code-switching. *Linguistics*, *18*(7/8), 581–618.

Poulisse, N. (1998). Review on: Duelling languages. Grammatical structure in codeswitching. *International Journal of Bilingualism*, *2*(3), 377–380. https://doi.org/10.1177/136700699800200308

Radford, A. (2009). *An introduction to English sentence structure*. Cambridge: Cambridge University Press.

Rey-Debove, J. (Ed.) (2014). *Le Petit Robert: Dictionnaire alphabétique et analogique de la langue française*. Nouvelle édition millésime 2014. Paris: Le Robert.

Rowlett, P. (2007). *The syntax of French*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511618642

Schachter, J. (1974). An error in error analysis. *Language Learning*, *24*(2), 205–214. https://doi.org/10.1111/j.1467-1770.1974.tb00502.x

Schmitt, E. (2000). Overt and covert codeswitching in immigrant children from Russia. *International Journal of Bilingualism*, *4*(1), 9–28. https://doi.org/10.1177/13670069000040010201

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*(1), 40–72.

Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209–231. https://doi.org/10.1515/iral.1972.10.1-4.209

Stevenson, A. (2010). *The Oxford dictionary of English* (3rd ed.). Oxford: Oxford University Press.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (3: Grammatical categories and the lexicon, pp. 57–149). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics. Vol. 2: Typology and process in concept.* Cambridge, MA: MIT Press.

Thomas, M. (2018). History of the study of second language acquisition. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 26–45). Cambridge: Cambridge University Press.

Timm, L. A. (1975). Spanish-English code-switching: El porqué y how-not-to. *Romance Philology*, *28*(4), 473–482.

Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research*, *12*(1), 7–39.

VanPatten, B., & Williams, J. (2015). Early theories in SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 17–33). New York: Routledge.

Wei, L. (1994). *Chinese (L1)/English (TL) learner corpus.*

Wei, L. (1995). *Chinese and Japanese learners' interlanguage data.*

Wei, L. (1996a). Organizational principles behind codeswitching and interlanguage development in early adult second language acquisition. In J. Arnold, R. Blake, B. Davidson, S. Schwenter, & J. Solomon (Eds.), *Sociolinguistic variation: Data, theory, and analysis* (pp. 417–431). Stanford: CSLI Publications.

Wei, L. (1996b). *Variation in the acquisition of morpheme types in the interlanguage of Chinese and Japanese learners of English as a second language* (Unpublished doctoral dissertation). University of South Carolina, Columbia, SC.

Wei, L. (2000a). Types of morphemes and their implications for second language morpheme acquisition. *International Journal of Bilingualism*, *4*(1), 29–43. https://doi.org/10.1177/13670069000040010301

Wei, L. (2000b). Unequal election of morphemes in adult second language acquisition. *Applied Linguistics*, *21*(1), 106–140. https://doi.org/10.1093/applin/21.1.106

Wei, L. (2003). An implicational hierarchy of morpheme acquisition order in second language learning. *ITL - International Journal of Applied Linguistics*, *139/140*, 77–100.

Wei, L. (2006). The multilingual mental lexicon and lemma transfer in third language learning. *International Journal of Multilingualism*, *3*(2), 88–104. https://doi.org/10.1080/14790710608668390

Wei, L. (2009). The composite nature of interlanguage as a developing system. *Research in Language*, *7*(1), 5–30. https://doi.org/10.2478/v10015-009-0002-9

Wei, L. (2015). *Interlanguage: The abstract level in language acquisition.* Lewiston: The Edwin Mellen Press.

Wei, L. (2018). Abstract lexical structure in second language learning. *Studies in Linguistics and Literature*, *2*(3), 223-243. https://doi.org/10.22158/sll.v2n3p223

Wei, L., & Liu, X. (2017). The bilingual mental lexicon and language transfer in second language learning. *International Journal of Language and Linguistics*, *4*(4), 11–23.

White, L. (1985). The "pro-drop parameter" in adult second language acquisition. *Language Learning*, *35*, 47–61.

Appendix

Table A1: Learner information and available data from the *Newcastle Corpus* (Myles & Mitchell, 2013c)

			Year 12 (recording round 1)						
Participant	Sex	Age	Grade	LT	MT	QT	NA	Ph	PD
8	female	16;9	A+	✓	✓	✓	✓	✓	✓
19	female	17;0	A	✓	✗	✗	✓	✓	✗
20	female	16;9	A	✓	✓	✓	✓	✗	✗
21	female	16;11	A*	✓	✓	✓	✓	✓	✓
23	female	16;8	A	✓	✓	✓	✓	✓	✓
24	male	16;11	A*	✓	✓	✓	✓	✓	✓
25	female	17;2	A*	✗	✓	✓	✓	✓	✓
27	female	17;4	A	✗	✗	✓	✓	✓	✓
28	male	16;7	A	✗	✓	✗	✓	✓	✓
30	male	17;4	A	✗	✓	✓	✓	✗	✓

Note. Age corresponds to the participants' age at the time of recording the Loch Ness task. The following abbreviations are used: LT = Loch Ness task, MT = Modern Times task, QT = Interrogatives task, NA = Negatives and Adverbs task, Ph = Photos task, PD = Pair Discussion task.

Table A2: Definite, indefinite, and partitive articles in French

Number/gender	**Definite article**	**Indefinite article**	**Partitive article**
SG.M	*le, l'*	*un*	*de, du* (*de* + *le*), *de l'*
SG.F	*la, l'*	*une*	*de, de la, de l'*
PL	*les*	*des*	*de, des* (*de* + *les*)

Note. If the word following the article has an initial vowel or silent *h*, the vowels of the singular forms of the definite and partitive article are elided: *le*/*la* > *l'*, *du*/*de la* > *de l'*.

Table A3: Strong and clitic pronouns in French (adapted from Gabriel et al., 2018, p. 155)

	Strong		**Clitic**		
Person/number	**Subject**	**Direct object**	**Subject**	**Direct object**	**Reflexive**
1SG	*moi*	*moi*	*je*	*me*	*me*
2SG	*toi*	*toi*	*tu*	*te*	*te*
3SG.M/F	*lui/elle*	*lui/elle*	*il/elle*	*le/la*	*se*
1PL	*nous*	*nous*	*nous*	*nous*	*nous*
2PL	*vous*	*vous*	*vous*	*vous*	*vous*
3PL.M/F	*eux/elles*	*eux/elles*	*ils/elles*	*leur*	*se*

Note. The paradigm for the indirect object pronouns has been left out here because it is not relevant for the discussion of the examples.

Table A4: Formation of subject and object *wh*-questions with the question marker *est-ce qui/est-ce que* in French

	Subject question ***est-ce qui***	**Object question** ***est-ce que***
***Qui* 'who'**	*Qui est-ce qui est dans la salle*?	*Qui est-ce que tu vois*?
	Who Q.SBJ is in the room	Who Q.OBJ you see
	'Who is in the room?'	'Who do you see?'
***Que* 'what'**	*Qu' est-ce qui est dans la salle*?	*Qu' est-ce que tu vois*?
	What Q.SBJ is in the room	What Q.OBJ you see
	'What is in the room?'	'What do you see?'

Note. *Que* is abbreviated to *qu'* if followed by a vowel-initial element.

List of abbreviations

2SG	second-person singular
3SG	third-person singular
1PL	first-person plural
2PL	second-person plural
3PL	third-person plural
C	complementizer
CLIT	clitic
CML	composite matrix language
CP	complementizer phrase
DEF	definite
DP	determiner phrase
EL	embedded language
F	feminine
FLLOC	*French Learner Language Oral Corpora*
IL	interlanguage
INF	infinitive
INFL	inflection
IP	inflection phrase
L1	native language
L2	second language
LT	"Loch Ness" task
M	masculine
MLF	matrix language frame
MT	"Modern Times" task
NA	"Negatives and Adverbs" task
NEG	negator
NP	noun phrase
partART	partitive article
PPTCP	past participle
PD	pair discussion task
PL	plural
PP	prepositional phrase
PT	"Photos" task
Q	question marker
Q.OBJ	object-question marker
Q.SBJ	subject-question marker
qc.	(French) *quelque chose* 'something'
qn.	(French) *quelqu'un* 'somebody'
QT	"Interrogatives" task
RQ	research question
SG	singular
SLA	second-language acquisition
TL	target language
VP	verb phrase