University of Pannonia, Veszprém
Faculty of Humanities
Multilingualism Doctoral School

PhD Dissertation

Hungarian-English Code-Switching:

A comparative analysis of code-switching frameworks on morphosyntactic integration



Written by: Salamon Attila

Supervisor: Prof. Margaret Deuchar

Veszprém

2026

DOI:10.18136/PE.2026.1019

Hungarian-English Code-Switching:
A comparative analysis of code-switching frameworks on morphosyntactic integration

Thesis for obtaining a PhD degree in the Multilingualism Doctoral School

of the University of Pannonia

in the field of Applied Linguistics

Written by Attila Salamon

Supervisor: Prof. Margaret Deuchar

Propose acceptance (yes / no)

…….

(Supervisor)

As a reviewer, I propose acceptance of the thesis:

Name of Reviewer: …… (yes / no) …….

(Reviewer)

Name of Reviewer: …… (yes / no) …….

(Reviewer)

The PhD candidate has achieved ……% at the public discussion.

Veszprém, …… / …… / 2026

…….

(Chair of the Committee)

The grade of the PhD Diploma …… (……%)

Veszprém, …… / …… / 2026

…….

(Chair of UDHC)

Abstract

Code-switching (CS) is generally understood as the use of more than one language in the same speech event, utterance, sentence or discourse. This study aims to demonstrate how naturally occurring Hungarian-English code-switching is produced, with special regard to morphosyntactic integration. The treatment of English-origin material in this language pair is compared under the Matrix Language Frame (MLF) Model and the Poplackian variationist approach, including the Nonce Borrowing Hypothesis (NBH). The study also aims to promote the use of publicly available podcast data for building linguistic corpus. The analysis draws on approximately 30 hours of naturally occurring speech from 24 Hungarian podcast videos involving 37 speakers. The corpus contains 419 bilingual constructions identified in 407 utterances, of which 405 intraclausal constructions were retained for structural comparison. The predictions of the two frameworks are tested based on the grammatical outcomes of the integration of English-origin material.

The results show extensive and systematic Hungarian grammatical control. Of the 405 intraclausal constructions, 187 involve intraword integration, including 152 lone English-origin items and 35 multiword expressions. English-origin bases received Hungarian case, number, possessive, verbal, adjectival, comparative, and nominalizing morphology and were placed in Hungarian coverb constructions. All 40 English-origin verbal constructions were formed by Hungarian verbalizer morphemes before possible further derivation or inflection.

The MLF Model provided an available structural analysis for all 405 intraclausal constructions: Hungarian consistently provided the morphosyntactic frame, and no unambiguous violation of the Morpheme Order Principle or the relevant system-morpheme predictions was identified. The Poplackian approach provided analysis for 370 constructions but left the 35 morphologically integrated multiword expressions indeterminate. As the frameworks mainly converge on the observable Hungarian structure while differing in the interpretation of integrated English-origin material as code-switching or borrowing, the conclusion that morphosyntactic integration can be established independently of contact-status classification is drawn. The study also highlights and demonstrates the value of podcasts as recheckable sources of researcher-unelicited bilingual speech, while recognizing their public, possibly performed, and non-representative characteristics.

Keywords: bilingualism, code-switching, intraclausal switching, intraword switching, morphosyntax, linguistic integration, podcast data, naturalistic speech, Matrix Language Frame Model, Nonce Borrowing Hypothesis, linguistic framework comparison

Acknowledgements

First and foremost, I would like to express my deepest gratitude towards Prof. Margaret Deuchar and Prof. Judit Navracsics for assisting me throughout my studies with patience, empathy and invaluable mental and academic support. Their guidance, honest feedback and motivation was essential in my development and the writing process of this dissertation. It was an honour to work and study under their supervision.

My sincere thanks to my wonderful colleagues at the English and American Studies Institute, who are always there to provide support with an open ear or collaborate with for any academic project.

Finally, I am most grateful to my loving wife, Alexandra Gyimesi, who has always been encouraging, supporting and as caring as a human can possibly be.

List of abbreviations

BIA+ – Bilingual Interactive Activation Plus model

CanVEC – Canberra Vietnamese-English Code-switching Natural Speech Corpus

CHILDES – Child Language Data Exchange System

CP – Complementiser Phrase

CS – Code-switching

EL – Embedded Language

Ex. – Example

Gl. – Glossing

L1 – First language

L2 – Second language

L^D – Donor Language

L^R – Recipient Language

LOLI – Lone Other-Language Item

ML – Matrix Language

MLID – Matrix Language Identity

MLF – Matrix Language Frame model

MOP – Morpheme Order Principle

NBH – Nonce Borrowing Hypothesis

SMP – System-Morpheme Principle

SVO – Subject-Verb-Object (word order)

Tr. – English translation

List of tables

Table 1. Examples of English-origin lexical material in Hungarian utterances from the data

Table 2. Examples of English verbal material integrated into Hungarian from Kontra (1995)

Table 3. Examples of multiword English material integrated into Hungarian from the data

Table 4. Examples of Hungarian nominal morphology applied on English stems

Table 5. Examples of English bases integrated by Hungarian verbalizers

Table 6. Examples of English-origin items receiving adjectival morphology

Table 7. Examples of integrated English items with Hungarian determiners

Table 8. Hungarian-English compounds from the data

Table 9. Bilingual examples with English system-morphology from the data

Table 10. Summary of the MLF and the Poplackian frameworks' predictions

Table 11. Multiword English material with Hungarian nominal morphology

Table 12. Examples of Hungarian verbalizer use on English items

Table 13. Adjectival integration in Hungarian morphosyntax

Table 14. Plural and case morphology applied on English-origin bases

Table 15. Hungarian coverbs combined with English-origin bases

Table 16. Lone English-origin items without overt morphology

Table 17. Compared framework coverage of the data

Table 18. Summary of the predictions and their outcomes

Table of Contents

1.Introduction **1**
1.1 Code-Switching as a Feature of Bilingual Speech 2
1.2 The Problem of Morphosyntactic Integration 2
1.3 The Theoretical Problem 3
1.4 Aims of the Dissertation 4
1.5 Research Questions 4
1.6 Significance of the Study 5
1.7 Overview of the Dissertation 6
2. Code-switching and Bilingual Language Production **7**
2.1 Defining Code-Switching 7
2.2 Bilingual Language Production 10
2.3 Translanguaging in Relation to Code-Switching 14
2.4 Grammatical Approaches to Code-Switching 15
2.5 Corpus-Based and Naturalistic Research on Code-Switching 17
2.6 Hungarian-English Language Contact and Code-Switching 19
2.6.1 Sociolinguistic Background of Hungarian-English Contact 20
2.6.2 Previous Research on Hungarian-English Code-Switching 21
2.6.3 Morphological Integration of English Elements in Hungarian 22
2.6.4 Research Gaps 23
3. Morphosyntactic Integration **25**
3.1 Descriptions of Morphosyntactic Integration 25
3.2 Morphosyntactic Integration in Comparable Contact Settings 27
3.2.1 Finnish-English Contact 27
3.2.2 Turkish Contact Morphology 28
3.2.3 Welsh-English Contact 28
3.2.4 Additional Evidence from Estonian and Croatian 29
3.3 Structural Properties of Hungarian Relevant to Code-Switching 30
3.3.1 Word Order 30
3.3.2 Nominal Morphology 31
3.3.3 Verbal Morphology and Verbalization 32
3.3.4 Adjectival Morphology and Recategorization 33
3.3.5 Coverbs and Verbal Prefixes 34

3.3.6 Determiners and Nominal Constructions 35
3.3.7 Compounding and Mixed Compounds 36
3.4 Structural Configurations of English-Origin Material 37
4. Competing Accounts of Bilingual Morphosyntax 39
4.1 Rationale for the Comparison of Theoretical Frameworks 39
4.2 The Matrix Language Frame Model 40
4.2.1 Development and Scope of the MLF Model 40
4.2.2 The Matrix Language and Embedded Language 41
4.2.3 The Morpheme Order Principle 41
4.2.4 The System Morpheme Principle 42
4.2.5 The 4-M Model 42
4.2.6 Embedded Language Islands 43
4.2.7 Congruence and the Blocking Hypothesis 44
4.2.8 Later Developments and Limitations of the MLF Model 44
4.3 The Poplackian Variationist Approach 45
4.3.1 The Equivalence Constraint 46
4.3.2 The Free Morpheme Constraint 46
4.3.3 Code-Switching and Borrowing 47
4.3.4 Established and Nonce Borrowing 48
4.3.5 The Nonce Borrowing Hypothesis 48
4.3.6 The Comparative Method and Recipient-Language Benchmarking 49
4.3.7 Criticisms and Methodological Limitations 50
4.4 Comparison of the Matrix Language Frame and Poplackian Variationist Approaches . 51
4.4.1 Shared Assumptions and Areas of Convergence 51
4.4.2 Matrix Language, Recipient Language, and the Interpretation of Integration 52
4.4.3 The Status of Lone and Internally Structured Other-Language Material 52
4.4.4 Frequency, Establishment, and Community Evidence 53
4.4.5 Evidential Requirements, Falsifiability, and Model Comparison 54
4.5 Predictions for Hungarian-English Morphosyntactic Integration 55
4.5.1 Predictions of the Matrix Language Frame Model 55
4.5.2 Predictions of the Poplackian Approach and the Nonce Borrowing Hypothesis 56
4.5.3 Summary of Framework-Specific Predictions 57
5. Data and Methods 59
5.1 Research Design 59
5.2 Data Source and Corpus 59

5.3 Podcast Selection Criteria 60
5.4 Identification of Bilingual Constructions and Units of Analysis 61
5.5 Transcription and Glossing 62
5.6 Coding and Analytical Classification 63
5.7 Quantitative Analysis 64
6. Results 65
6.1 Overview of the Corpus 65
6.2 Morphosyntactic Integration Patterns 65
6.2.1 Nominal Integration 66
6.2.2 Verbal Integration 67
6.2.3 Adjectival Integration 68
6.2.4 Case and Number Marking 68
6.2.5 Possessive Morphology 69
6.2.6 Coverbs 70
6.2.7 Mixed Compounds 70
6.2.8 Multiword English Material and Potential EL Islands 71
6.3 Recategorization 72
6.4 Testing the MLF Predictions 73
6.4.1 Matrix Language Identification 73
6.4.2 The Morpheme Order Principle 74
6.4.3 The System Morpheme Principle and 4-M Predictions 75
6.4.4 Embedded Language Islands 76
6.4.5 Problematic Cases 76
6.5 Testing the Poplackian Predictions 77
6.5.1 Lone English-Origin Items 77
6.5.2 Morphosyntactic Integration 78
6.5.3 Established and Nonce Forms 78
6.5.4 Multiword Switches 79
6.5.5 Problematic and Indeterminate Cases 79
6.6 Comparative Framework Coverage 81
7. Discussion 82
7.1 Summary of the Major Findings 82
7.2 Empirical Evaluation of the MLF Model 83
7.3 Empirical Evaluation of the Poplackian Approach 85
7.4 Where the Frameworks Converge 88

7.5 Where the Frameworks Diverge 89
7.6 Morphological Integration Without Theoretical Agreement 92
7.7 Hungarian Morphology as a Test Case for Code-Switching Theory 95
7.8 Morphophonological Integration as Supporting Evidence 97
7.9 Code-Switching Versus Borrowing Reconsidered 99
7.10 Implications for Models of Code-Switching 100
7.11 Methodological Implications of Podcast Data 102
7.11.1 Naturalistic Bilingual Interaction 102
7.11.2 Recheckability and Analytical Transparency 103
7.11.3 Naturalistic Speech as Performed Discourse 104
7.11.4 Sampling and Speaker Information 104
7.11.5 Implications for Corpus Interpretation 105
7.12 Limitations 106
8. Conclusions 108
8.1 Answers to the Research Questions 108
8.2 Main Theoretical Conclusions 109
8.3 Contributions of the Dissertation 110
8.3.1 Empirical Contribution 110
8.3.2 Theoretical Contribution 111
8.3.3 Methodological Contribution 111
8.4 Directions for Future Research 111
References 113
Appendices 121

1. INTRODUCTION

Speakers of more than one language have always been in the focus of linguistic research; however, different methods, frameworks and language pairs might yield different results, therefore diverging conclusions can be drawn from such language-contact studies. This dissertation is written with the aim of testing different frameworks applied on the same corpus in an objective and scientifically sound manner, while analysing how other-language material is being incorporated during spontaneous speech.

Hungarian-English bilingual speech frequently combines English lexical material with Hungarian grammatical structure. In the podcast corpus examined in this dissertation, speakers produced forms such as *content-et* 'content-ACC', *stream-el-ünk* 'stream-VERB-1PL', and *fake meat-ek-ért* 'fake meat-PL-CAUS-FIN'. These examples contain recognizably English lexical material, yet their case marking, category assignment, number, inflection, and morpheme order are supplied by Hungarian (see Table 1).

Table 1. Examples of English-origin lexical material in Hungarian utterances from the data

Ex.1	*"a gaming content-et hiányol-od?" (*tr.: *do you miss the gaming content?)*
Gl.:	the.DET.DEF gaming.N content.N-ACC miss.V-2S.ACC
Ex.2	*"mi stream-el-ünk"* (tr.: *we stream*)
Gl.:	we.PRON.PERS stream.N-VBZ-1PL
Ex.3	*"nem ugrándoz-om ez-ek-ért a fake meat-ek-ért"* (tr.: *I'm not fond of these fake meats*)
Gl.:	no.NEG jump.V-1S these.PRON.DEM.PL-CAUS-FIN the.DET.DEF fake.ADJ meat.N-PL-CAUS-FIN

Such constructions raise questions that cannot be answered by identifying the languages in the utterance alone. They require an explanation of how material from two grammatical systems is combined and whether the resulting form should be analysed as code-switching [CS], borrowing, or a construction with a contact status that cannot be determined from its surface structure.

1.1 Code-Switching as a Feature of Bilingual Speech

Code-switching is generally understood as the use of more than one language within the same interaction, utterance, clause, or constituent (Bullock & Toribio, 2009). It is a systematic feature of bilingual speech rather than linguistic confusion or incomplete multilingual competence (Grosjean, 1989). Bilingual speakers may alternate between longer stretches of their languages, insert lexical elements, or construct forms in which material from both languages occurs within a single word (Gardner-Chloros, 2009; Poplack, 1980).

The present dissertation is concerned primarily with intraclausal and intraword combinations. The data are not extended alternations between complete Hungarian and English clauses, but constructions in which English-origin material participates in Hungarian syntax and morphology. The focus is less on why a speaker chooses an English expression in discourse and more on how that expression is grammatically incorporated.

The term *English-origin item* is used where an initially theory-neutral description is required, so as to avoid assuming in advance that every recognisably English form is necessarily a code-switch – even though the author might apply that term in future observations, as the title suggests. Framework-specific classifications are applied only after the construction's morphological and syntactic properties have been established.

1.2 The Problem of Morphosyntactic Integration

The main empirical issue of the study is what happens when an English lexical item enters a Hungarian morphosyntactic environment. Hungarian is a morphologically rich language in which grammatical relations are frequently expressed through affixation. English-origin nouns can receive Hungarian case, plural, or possessive morphology, while some English-origin bases can be converted into Hungarian verbs or adjectives before receiving further inflection.

A form such as *content-et* is not simply an English word placed beside Hungarian words. The accusative relation is encoded through a Hungarian suffix with a form compatible with Hungarian morphophonology. Similarly, *stream-el-ünk* contains an English lexical base, a Hungarian verbalizer, and Hungarian first-person plural inflection. Therefore, the English element is incorporated into a structure that determines its syntactic category, grammatical function, and morphological realization.

These constructions make the contribution of Hungarian data valid and clearly visible. However, overt integration does not automatically determine the theoretical status of the English-origin material. The same morphology may be interpreted differently depending on the applied model.

1.3 The Theoretical Problem

The dissertation compares two influential approaches to bilingual speech: the Matrix Language Frame [MLF] Model (Myers-Scotton, 1993, 2002) and the Poplackian variationist approach including the Nonce Borrowing Hypothesis [NBH] (Poplack, 1980, 2012, 2018; Sankoff et al., 1990).

According to the MLF Model, bilingual clauses are structurally asymmetrical. One language supplies the morphosyntactic frame, while the other contributes Embedded Language [EL] material. English lexical elements can receive Hungarian morphology and be analysed as code-switches. On this account, forms such as the examples of Table 1 demonstrate the control of the Hungarian Matrix Language [ML] over English Embedded Language material. This asymmetrical organization is central to this model (Myers-Scotton, 1993, 2002).

The Poplackian approach emphasises the distinction of code-switching from borrowing through grammatical behaviour. Lone donor-language [L^D] items that pattern with the recipient language [L^R] are analysed as established or nonce borrowings rather than as code-switches. While morphosyntactic integration is central to their classification, phonological integration is less decisive because these processes do not necessarily occur and develop together. The possibility that a socially unestablished item can immediately display recipient-language grammar is formalized in the Nonce Borrowing Hypothesis (Poplack, 2012, 2018; Sankoff et al., 1990).

The disagreement of the models is not primarily about whether the Hungarian morphology is present, as both approaches recognize that the English-origin material participates in Hungarian structure. They differ in their theoretical interpretations: for the MLF, Hungarian morphology can be part of an asymmetrical code-switching construction, whereas for the Poplackian approach, recipient-language-equivalent behaviour supports the case of a borrowing.

The Hungarian-English corpus is particularly useful for investigating this disagreement as Hungarian morphology provides several overt diagnostics of grammatical integration, and it

makes it possible to compare the frameworks on the same naturally occurring constructions rather than on examples drawn from separate studies or language pairs. Deuchar (2020) identifies the distinction between code-switching and borrowing and the empirical evaluation of grammatical accounts as central issues in contemporary code-switching research and advises not assigning all lone other-language items to a single category based on one integration criterion.

1.4 Aims of the Dissertation

The primary aim of this dissertation is to describe how English-origin material is integrated into Hungarian morphosyntax in naturally occurring podcast speech. The second aim is to evaluate the extent to which the MLF Model accounts for the structural organization of the resulting bilingual constructions, with particular attention to ML identification, morpheme order, system morphology, category assignment, and the treatment of complex English material. Third, the study examines the same data from the Poplackian perspective, focusing on the distinction between integrated lone items, borrowing, nonce borrowing, and multiword code-switches. Finally, to compare the approaches and where they converge or diverge in their description of the data, classification, and which constructions remain theoretically indeterminate.

1.5 Research Questions and Predictions

Considering the aims, the dissertation addresses the following research questions:

1. How is English-origin material integrated into Hungarian morphosyntax in naturally occurring bilingual speech?
2. To what extent can the intraclausal constructions be accounted for by the Matrix Language Frame Model?
3. To what extent are the data compatible with the Poplackian approach and the Nonce Borrowing Hypothesis?
4. Where do the frameworks converge and diverge in their treatment of the same constructions?

Several predictions follow from considering the theoretical literature. First, if Hungarian functions as the Matrix Language, it should determine the principal morphosyntactic frame of the mixed clauses. English-origin material is expected to follow Hungarian morpheme order and to receive Hungarian category-forming, inflectional, and system morphology if necessary.

Violations of the Morpheme Order Principle [MOP] or the relevant predictions of the System-Morpheme Principle [SMP] (see Chapter 4) are not expected.

Second, according to the Poplackian approach, lone English-origin items that display Hungarian grammatical behaviour should be compatible with a borrowing or nonce-borrowing interpretation. Longer English sequences that retain identifiable English internal grammar should be classified as code-switches.

Third, constituent size is not expected to produce a completely categorical distinction. Multiword English material may form a compound or complex lexical base and may receive Hungarian morphology externally. Such constructions may be less easily classified within the canonical opposition between lone borrowing and multiword switching.

Finally, morphophonological integration is expected to serve as evidence of Hungarian grammatical control but not in determining borrowing status independently. Hungarian vowel harmony, suffix selection, and phonological accommodation may accompany morphosyntactic integration without providing a categorical boundary between code-switching and borrowing (Stefanich et al., 2019).

1.6 Significance of the Study

The dissertation offers an empirical contribution by providing a detailed analysis of Hungarian-English bilingual constructions in approximately 30 hours of podcast speech, since this language pair is less extensively described than other major bilingual language pairs (e.g. Spanish-English), especially with respect to word-internal and multiword morphological integration.

The theoretical significance lies in the direct comparison of competing analyses on the same dataset. This design avoids evaluating MLF and the Poplackian approach on different forms or different language pairs and intends to draw a bridge between theory and data by distinguishing disagreements about empirical structure from disagreements on theoretical classification.

The study carries methodological significance as well. Podcast recordings provide extended bilingual interaction that was not elicited for linguistic research. They preserve the constructions in their discourse contexts and allow reproducible examination of morphological, syntactic, and morphophonological interactions of the languages at hand. Although podcast

speech is public and potentially performed, it provides strong positive evidence of linguistic configurations speakers can produce without any researcher prompting.

The broader significance of the dissertation lies in separating two questions: how bilingual material is grammatically structured and how it should be classified within language-contact theory. Hungarian morphosyntax often makes the answer to the first question relatively clear, while the latter remains dependent on the assumptions and evidential requirements of the given framework. This distinction gives the focus for the following chapters.

1.7 Overview of the Dissertation

In the following chapters, previous work on language contact phenomena (code-switching and bilingual language production), morphosyntactic integration and more specifically the two frameworks at hand are considered. Following that, the methodology and the nature of the data is discussed in detail. After the presentation of the quantitative and qualitative results of the study, their interpretation is considered based on all the predictions the frameworks imply. Lastly, the Conclusion chapter contains the main contributions and limitations of the study as well as the possibilities for future research.

2. CODE-SWITCHING AND BILINGUAL LANGUAGE PRODUCTION

2.1 Defining Code-Switching

Code-switching is one of the most noticeable manifestations of bilingual language use, yet its general description as the alternation between two languages covers considerable terminological and theoretical complexity. In its broadest sense, it refers to the use of two or more languages within the same interaction, discourse, sentence, or utterance. Research on bilingual speech has demonstrated that such alternation is not arbitrary or indicative of linguistic deficiency but is a systematic feature governed by grammatical, psycholinguistic, sociolinguistic, and interactional factors (Auer, 1998; Bullock & Toribio, 2009; Gardner-Chloros, 2009; Gumperz, 1982; Milroy & Muysken, 1995; Muysken, 2000). Consequently, contemporary approaches generally treat code-switching as part of bilingual linguistic competence. Code-mixing, language alternation and code-switching are often used as synonymous terms, although most of the literature discussed in this study applies the latter one.

Foundational research on bilingualism also warns against evaluating bilingual language use as the imperfect combination of two monolingual systems. Grosjean (1982, 1989) supports a holistic view according to which the bilingual is a specific and integrated speaker – hearer whose linguistic capabilities reflect the coexistence and interaction of more than one language. According to this view, bilingual performance should not be measured against the expectation that each language will behave exactly as it does in a monolingual speaker. From a Hungarian psycholinguistic perspective, Navracsics (2010) similarly treats code-switching and code-mixing in spontaneous speech as manifestations of bilingual language processing that may occur at several structural levels, including morpheme boundaries, individual lexical items, and larger phraseological units.

The recognition of the systematic nature of code-switching was particularly important in early grammatical research. Poplack's (1980) analysis of Spanish-English bilingual speech demonstrated that switches were not distributed randomly throughout bilingual utterances. Instead, switches showed recurrent structural patterns, inducing the formulation of grammatical constraints intended to explain where switching is expected to occur. This work was later polished by Sankoff and Poplack (1981), who proposed constraints on intra-sentential switching and contributed to establishing code-switching as a legitimate object of grammatical

analysis. Although some later approaches differ in their theoretical assumptions, they share the recognition of bilingual utterances exhibiting systematic morphosyntactic organization (Myers-Scotton, 1993, 2002; Muysken, 2000).

An important distinction is frequently made in the literature concerning the location of language alternation. Intersentential switching occurs across sentence boundaries, which allows relatively longer stretches of speech in each language to occur. In contrast, intrasentential or intraclausal switching involves material from different languages within the same sentence or clause. The difference is that intrasentential switching deals with larger segments, so it can contain intraclausal switches (Deuchar, 2012). Using clauses as units of analysis is particularly important for grammatical research because the linguistic systems involved often interact within a single syntactic structure. Poplack (1980), Myers-Scotton (1993), and Muysken (2000), among others, have paid attention to switching within grammatical units. For the purposes of this study, such cases are especially informative because English-origin lexical material frequently occurs inside otherwise Hungarian morphosyntactic structures in the data.

The structural diversity of bilingual speech makes it problematic to consider all instances of language alternation as manifestations of a single process. Muysken (2000), for example, distinguishes three major patterns of code-mixing: insertion, alternation, and congruent lexicalization. Insertion is the incorporation of material associated with one language into a structure of another language; alternation involves switching between relatively monolingual stretches of different grammatical systems; and congruent lexicalization refers to structures where the languages share sufficient grammatical organization for lexical material from both to occur in a common frame. This distinction is somewhat relevant to Hungarian-English bilingual speech because many constructions involve an English lexical element embedded within morphologically and syntactically Hungarian material. Although describing such constructions as insertions would not settle the theoretical question of whether English-origin elements should be classified as code-switches or as borrowings.

This problem becomes evident in the treatment of lone other-language items [LOLIs]. Multiword English material retaining donor-language internal grammar can often be identified as code-switching, whereas classifying lone English-origin items as insertion or alternation depends on the construction and the applied framework. An isolated English-origin lexical item carrying Hungarian morphological marking is considerably more ambiguous. Such an item may be interpreted as an established borrowing, an occasional or nonce borrowing, a single-word

code-switch, or other bilingual forms depending on the theoretical criteria applied. The distinction is not just terminological, as competing approaches assign different theoretical significance to morphological and syntactic integration, which makes the classification of LOLIs a key controversy in research on bilingual speech.

The discussion on the relationship between code-switching and borrowing has been in a central position in language-contact research since the foundational work of Haugen (1950) and Weinreich (1953). According to them, borrowing is generally associated with the incorporation of material originating in one language into another, however, determining when such incorporation takes place is much more difficult than identifying the historical origin of a lexical form. Frequency, conventionalization, phonological adaptation, morphological integration, syntactic behaviour, community distribution, and dictionary attestation have all been used as potential criteria. However, classification is difficult as theoretical approaches differ in the emphasis attributed to each of these criteria.

This distinction is especially prominent in the variationist tradition associated with Poplack and colleagues. The works of Poplack and Sankoff (1984), Poplack et al. (1988), Sankoff et al. (1990), and Poplack (2018) distinguish established loanwords from occasional borrowing while arguing that the grammatical behaviour of donor-language-origin material within the recipient language provides important evidence regarding the process. Poplack's (2018) later synthesis places particular emphasis on morphosyntactic integration and on the comparison of other-language items with the grammatical patterns of the recipient language. Her approach also emphasizes the importance of corpus evidence capable of distinguishing individual or occasional (nonce) usages from broader community patterns.

An important claim of her work is that phonological integration is not a reliable criterion on its own. Poplack (2018) reports considerable variability in the phonetic realization of other-language material and argues that phonetic and morphosyntactic integration operate independently. So, according to her, morphosyntactic behaviour provides a more consistent diagnostic for distinguishing the different types of language mixing than phonetic adaptation. This claim is especially relevant to the Hungarian-English contact because English-origin lexical material may undergo Hungarian phonological accommodation, including vowel-harmonic suffix selection (see Section 7.8), while its status as a borrowing or a code-switch remains theoretically contestable.

Also, this study applies terminology with caution during the discussion of description and classification. Using the term *code-switch* for every English lexical element identified in mainly Hungarian discourse would imply a theoretical position before the competing frameworks have been evaluated. Conversely, treating every morphologically integrated English element as a borrowing would presuppose the validity of the criteria that is part of the theoretical question at hand. Thus, the more neutral expressions *English-origin item* and *English-origin material* are used where the descriptive status of the lexical material is clear but its classification as borrowing or code-switching remains theoretically open. Similarly, the term *intraword mixed form* may be used descriptively for words containing English-origin lexical material with Hungarian morphemes without assuming that such forms are instances of code-switching in both frameworks.

While discussing the general linguistic definition of code-switching, it should be understood as part of a broader range of bilingual language-contact phenomena in which linguistic resources conventionally associated with different languages become available within the same communicative event; whereas the present dissertation consequently treats the code-switching – borrowing distinction as an empirical and theoretical question to be tested rather than as a definitional assumption.

Before examining the grammatical mechanisms proposed to account for such integration, however, it is necessary to consider the psycholinguistic conditions that make bilingual language alternation possible.

2.2 Bilingual Language Production

The production of bilingual speech requires speakers to manage linguistic knowledge associated with more than one language while constructing an utterance. Although the observable result may involve only one language or a combination of two languages, bilingual language production does not involve one linguistic system being fully activated while the other remains completely unavailable. Psycholinguistic research instead indicates that bilingual processing involves interaction between linguistic representations associated with both languages (Dijkstra & van Heuven, 2002; Green, 1998; Kroll & Stewart, 1994). Therefore, lexical activation, language selection, inhibition, proficiency, and interactional context are directly relevant to understanding how an English-origin lexical item can become available during predominantly Hungarian speech and enter a Hungarian morphosyntactic structure.

Models of bilingual production partly build on general theories of speech production. Levelt's (1989) influential work distinguishes between conceptual preparation, formulation, phonological encoding, and articulation. Although originally developed as an account of monolingual speech production, his model provides an important stepping stone for considering the additional demands faced by bilingual speakers. At the conceptual level, the speaker develops the preverbal message to be expressed, then retrieves lexical and grammatical information and organises it before the final phonological representation is articulated. In bilingual speakers, however, lexical knowledge is distributed across their two linguistic systems, where lexical candidates of different languages become activated and selected.

The relationship between bilingual lexical representations and conceptual knowledge has been investigated by Kroll and Stewart (1994). Their Revised Hierarchical Model proposes that lexical and conceptual connections are affected by bilingual experience and proficiency. According to it, lexical links from L2 to L1 are relatively strong during earlier stages of L2 development, whereas increasing proficiency is associated with stronger direct links between L2 lexical and conceptual representations. Although subsequent work has refined this account, its general implication remains relevant: the accessibility of lexical items in the two languages is not necessarily equivalent and it might vary with proficiency, experience, frequency of use, and language dominance. Consequently, the selection of an English lexical item during Hungarian discourse cannot be understood only in terms of the existence of a Hungarian translation equivalent.

Research on bilingual lexical processing has also provided extensive evidence for language non-selective activation. Dijkstra and van Heuven's (2002) Bilingual Interactive Activation Plus model was developed mainly to account for bilingual word recognition. It assumes that lexical candidates from both languages become activated rather than access being restricted in advance to the language currently required. The model incorporates orthographic, phonological, and semantic representations and distinguishes processes involved in lexical identification from task- and decision-related mechanisms. Although the BIA+ is a word recognition model and so it should not be treated as a direct model of bilingual speech production, the broader evidence for cross-language activation is significant for challenging a strict language-selective view.

As a complementary perspective on this interaction, Navracsics (2007a, 2011) examines how lexical knowledge associated with two languages is organized and activated, emphasizing that bilingual lexical processing is influenced by several interacting factors, including language

proficiency, frequency of use, age of acquisition, and the relationship between the languages involved. According to her, the two languages cannot be represented as entirely independent lexical stores that are accessed without mutual influence. This is consistent with Grosjean's (1989) holistic view, according to which bilingual linguistic organization constitutes an integrated system shaped by the speaker's communicative history and language use.

If lexical candidates associated with both languages can become activated, speakers need mechanisms that allow appropriate items to be selected according to their communicative goals and contextual demands. Green's (1998) Inhibitory Control Model provides one of the most influential contributions to describing this process. The model proposes multiple levels of control and associates lexical selection with language task schemas capable of regulating competing representations. It is important to note that inhibition should not be understood as the complete deactivation of a language, but as a dynamic bilingual control which can operate over the competing items to different degrees depending on the demands of the task and the communicative situation.

Switching cost is a relevant and frequently studied area of bilingual speech production. Studies in which bilingual speakers are required to alternate between languages during naming tasks have identified measurable costs of switching relative to remaining in the same language. Meuter and Allport (1999) reported asymmetric switching costs, contributing to the proposal that inhibitory control plays a role in language selection. Costa and Santesteban (2004) later demonstrated that switching patterns are affected by bilingual proficiency, indicating that highly proficient bilinguals do not need to regulate their languages in the same way as less proficient L2 speakers. Their findings indicate that bilingual language selection is dynamic and sensitive to experience.

Experimental switching costs, however, must be distinguished from naturally occurring code-switching, as laboratory research typically instructs participants to use a given language according to an externally presented cue. Thus, the location and direction of the switch are possibly imposed by the experimental design. Natural bilingual interaction differs in this regard: speakers may determine whether and when they switch, and their choices may be influenced by a multitude of extralinguistic factors.

As natural bilingual language use can involve substantially different control demands from experimentally cued alternation, the bilingual production system must be considered in relation to the contexts in which it is used. Green and Abutalebi's (2013) Adaptive Control Hypothesis

develops this perspective by proposing that language-control processes adapt to the interactional demands. Different bilingual environments impose different requirements on language selection. Thus, the production of an English-origin item during an otherwise Hungarian discourse should not be interpreted as a failure of lexical retrieval or language control automatically. Han et al. (2022) found associations between habitual switching patterns and performance in verbal switching, nonverbal shifting, and inhibition tasks, which provides support for the Adaptive Control Hypothesis, and similarly, Özkara et al. (2025) emphasize that conclusions about code-switching and cognitive control depend on the operationalisation of switching experience, interactional context, and control processes.

Naturalistic production may yield different conclusions from externally cued switching tasks. Johns and Steuck (2021) found that speech rates were faster in code-switched discourse than in monolingual stretches in an established Spanish-English bilingual community, which suggests that code-switching may facilitate production at the discourse level even where localized processing costs are detectable.

Grosjean's (2001) concept of language mode provides a more specific account of how communicative context affects the relative activation of a bilingual's languages. It refers to the state of activation of the different languages and language-processing mechanisms at a particular point in time. Bilingual speakers may occupy different positions on a continuum extending from a more monolingual mode, in which one language is strongly activated and the other is less available, to a bilingual mode, in which both languages are activated and may contribute to the interaction. Although the speakers' position on this continuum cannot be measured in this study, the concept still helps explain why English lexical access can vary according to interlocutor, topic and shared linguistic knowledge. However, language activation may be more dynamic and less selectively compartmentalised than a straightforward interpretation of this continuum implies. Evidence for non-selective lexical access indicates that elements of both languages may remain active even when bilinguals are operating primarily in one language, although the extent of such cross-language activation is influenced by factors such as proficiency and linguistic context (van Hell & Tanner, 2012).

The distinction between lexical accessibility and grammatical organization becomes particularly important when English-origin elements receive Hungarian morphology. Navracsics's (2007b) examination of word classes in the bilingual mental lexicon is especially relevant to the relationship between lexical selection and grammatical integration. Lexical

representations involve not only phonological and semantic information but also grammatical category and combinatorial possibilities. However, in bilingual production, the category associated with a lexical item in one language does not necessarily determine how that item functions in a construction organized by the other language. An English-origin base selected during Hungarian production may therefore require Hungarian category-forming morphology before it can participate in the relevant structure. The bilingual mental lexicon makes English lexical material available, while Hungarian morphosyntax determines how that material is categorized and grammatically realized in the utterance.

2.3 Translanguaging in Relation to Code-Switching

An alternative perspective on bilingual language use is offered by translanguaging, a concept that developed from its original pedagogical use in bilingual education into a broader account of bilingual communicative practice (Lewis et al., 2012). In its stronger theoretical formulation, it questions the assumption that bilingual speakers operate with two neatly bounded linguistic systems. Otheguy et al. (2015) conceptualize bilingual linguistic knowledge as a structured but unitary repertoire of features and consequently distinguish translanguaging from code-switching: rather than selecting between two autonomous grammars, bilingual speakers select features from their available linguistic repertoire. Similarly, Li Wei (2018) presents translanguaging as a theory intended to capture linguistic practices that transcend the boundaries conventionally imposed between languages and extends the concept beyond strictly verbal behaviour to multilingual, multimodal, and multisemiotic meaning-making. In this respect, translanguaging differs from approaches such as the MLF Model and the Poplackian tradition, both of which retain language-specific categories in analysing bilingual constructions.

However, the theoretical status of these boundaries remains debated. While recognizing the value of the heteroglossic perspective associated with translanguaging, MacSwan (2017) argues that the sociopolitical character of named languages should not be taken to imply that linguistic differentiation is analytically or structurally meaningless; accordingly, his integrated multilingual model distinguishes a speaker's linguistic repertoire from the grammatical systems represented within it. Here, translanguaging is not operationalized as a third competing morphosyntactic framework, since doing so would alter the theoretical assumptions underlying the object of comparison itself. Nevertheless, it provides an important complementary perspective: the labels *Hungarian* and *English* used in the analysis should be understood as analytically necessary categories for testing the two selected frameworks, not necessarily as

evidence that bilingual speakers experience or use their linguistic resources as two completely autonomous systems.

2.4 Grammatical Approaches to Code-Switching

The grammatical study of code-switching has developed through a series of competing accounts of explanations of how material from different languages can be combined within a single utterance without violating the structural requirements of either language. Since it has been established that code-switching is not random, research interest revolves around the grammatical constraints that regulate mixed constructions. Different approaches have addressed this problem, and they vary in their assumptions concerning language-specific grammar, structural asymmetry, lexical insertion, constituent boundaries, and the status of morphologically integrated other-language material.

One of the earliest influential approaches was developed by Poplack (1980), whose analysis of Spanish-English bilingual speech proposed that intra-sentential switching was subject to grammatical restrictions. Two constraints became particularly influential: the Free Morpheme Constraint and the Equivalence Constraint. The Free Morpheme Constraint restricts switching at the boundary between a bound morpheme and a lexical stem. The Equivalence Constraint, in turn, proposed that switching tends to occur at points where the surface structures of the participating languages are compatible, thereby avoiding configurations in which a switch would violate the syntactic organization of either language. Subsequently, Sankoff and Poplack (1981) developed these ideas in a more formal treatment of bilingual sentence structure.

On the other hand, the notion of asymmetry became prominent in Myers-Scotton's (1993, 2002) MLF model, which proposes that the languages participating in intra-sentential code-switching do not necessarily make equivalent structural contributions. It formulates this asymmetry through specific principles governing morpheme order and the distribution of system morphemes. The model was later refined through the 4-M model developed by Myers-Scotton and Jake (2000, 2001). Rather than treating all grammatical morphemes as a single class, the 4-M model distinguishes content morphemes from several types of system morphemes according to their activation and relationship to conceptual structure. The distinction between early and late system morphemes, and between bridge and outsider late system morphemes, was intended to provide a more precise account of the distribution of grammatical material in bilingual speech. These developments are particularly relevant to

morphologically rich matrix languages (e.g. Hungarian) because they offer explicit predictions about which types of grammatical morphology should be expected to originate from these languages. The Poplackian approach and the MLF and 4-M models are discussed in greater detail in Chapter 4, where their predictions are evaluated for the Hungarian-English data.

Mentioned earlier, a different type of account is represented by Muysken's (2000) typology of bilingual speech. His distinction of insertion, alternation, and congruent lexicalization is important for two reasons. First, it emphasizes that different types of code-switching may require different grammatical explanations rather than being reducible to a single mechanism. Second, it provides a useful descriptive framework for distinguishing between lexical insertion and larger-scale alternation. Since the corpus predominantly consists of insertions, and the focus of this study is on intraclausal switches, Muysken's framework is used here primarily as a broader typological point of reference rather than as a model under systematic testing.

Rather than assuming that bilingual speech requires a separate set of code-switching-specific constraints, generative approaches seek to explain mixed constructions through general principles of grammatical theory. Di Sciullo et al. (1986), for example, proposed an analysis based on government relations, while Belazi et al. (1994) formulated the Functional Head Constraint. Also, the works of Mahootian (1993) and Mahootian and Santorini (1996) argue for approaches in which grammaticality comes from the ordinary syntactic properties of the languages involved in the CS process.

A similar position is taken in MacSwan's (1999, 2000, 2005) work, which is also developed in the contributions collected in MacSwan (2014). In his Minimalist account, he argues that code-switching does not require a separate set of code-switching-specific grammatical constraints; mixed sentences should instead be derived using independently motivated principles of grammar. Therefore, bilingual utterances are generated by the same general grammatical architecture that underlies monolingual language, with lexical items from more than one language available for selection during derivation. He therefore rejects the idea that code-switching requires an additional set of language-mixing constraints beyond the independently motivated principles of grammar.

This position led to a theoretical debate with the MLF approach. MacSwan (2005), criticized several aspects of the Matrix Language Frame Model, including its use of language-specific structural asymmetry and the explanatory status of the Matrix Language construct; however, Jake et al. (2002) argued that a successful account of bilingual syntax requires recognition of

the unequal structural roles played by the participating languages. This disagreement reflects a broader theoretical divide in code-switching research: one approach seeks to derive bilingual structures from general grammatical principles without a specialized code-switching mechanism, while the other argues that recurrent asymmetries in mixed speech require an explicit model of language dominance within the bilingual clause.

2.5 Corpus-Based and Naturalistic Research on Code-Switching

Since bilingual linguistic choices are sensitive to interlocutor, topic, communicative setting, language dominance, social relationships, and community norms, the conditions under which bilingual speech is recorded can influence both the frequency and the structural characteristics of switching. This has made naturally occurring conversational data particularly important in code-switching research.

The importance of naturally occurring data is closely connected with the broader sociolinguistic problem described by Labov (1972) as the observer's paradox: linguistic research seeks to find out how speakers use their language(s) when they are not being observed, yet the act of observation itself can influence linguistic behaviour. This problem is especially relevant to the current study. Speakers may reduce switching when interacting with a researcher whom they perceive as someone expecting monolingual language use, while an experimental task explicitly requiring language alternation may increase switching in ways that do not resemble ordinary interaction. Accordingly, the linguistic behaviour documented in controlled settings can differ from the patterns speakers produce when language choice remains under their own control.

Poplack's (1980) analysis of Puerto Rican Spanish-English bilingual interaction, for example, relied on naturally produced speech to identify recurrent switching patterns and formulate structural constraints. Also, Poplack et al. (1988) investigated lexical borrowing through quantitative analysis of naturally occurring bilingual discourse, while Poplack et al. (1989) used corpus evidence from Finnish-English bilinguals to distinguish different language-contact phenomena. These studies illustrate an important methodological principle: the classification of individual bilingual forms becomes more reliable when they can be compared with broader patterns of usage in the recipient language and speech community.

More recent corpus-building initiatives demonstrate the continued importance of natural speech data for code-switching research. For example, Nguyen and Bryant (2020) developed

CanVEC, a corpus of naturally occurring Vietnamese-English bilingual speech, specifically to support sociolinguistic and code-switching research.

Corpus-based research is also particularly valuable because code-switching is structurally heterogeneous. Different types of bilingual constructions may occur at very different frequencies, and many theoretically significant forms are difficult to elicit directly. Gullberg et al. (2009) emphasize that research on code-switching benefits from combining appropriate data-collection techniques with detailed linguistic annotation, since the research question determines which properties of bilingual speech must be documented. Naturalistic corpora allow the examination of switches in the actual environments where they occur, making them especially useful for analyses of grammatical integration.

Nevertheless, the construction and analysis of bilingual corpora have their methodological difficulties beyond those in monolingual corpus research. Deuchar et al. (2014) note that bilingual corpus building requires making decisions about language identification, transcription conventions, annotation, speaker metadata, and the representation of mixed constructions. These decisions are not theoretically neutral, as labelling isolated lexical items may be difficult when they are historically borrowed, phonologically ambiguous, structurally integrated, or shared across languages. Doğruöz et al. (2021) show that code-switching displays substantial structural and functional diversity across multilingual communities and argue that existing data often fail to represent the full range of language pairs, switching types, and sociolinguistic settings.

An additional methodological challenge is transcription and annotation. As Gullberg et al. (2009) emphasize, the transcriptional detail required in code-switching research depends on the analytical objectives. Morphosyntactic research requires reliable identification of lexical material and morphological boundaries, while work involving morphophonological integration may additionally require attention to pronunciation and phonological adaptation. Corpus standards developed within broader spoken-language research, including systems associated with CHILDES and TalkBank (MacWhinney, 2000), demonstrate the importance of transparent and replicable transcription and annotation conventions.

However, a major advantage of a naturalistic corpus is that it is well suited to morphosyntactic analysis because grammatical integration is only observable within an appropriate structural environment. While an isolated lexical item provides limited evidence about the grammar governing its use, corpus data allow the analysis of mixed items not simply

as lexical tokens but as components of larger grammatical constructions. Besides that, quantitative corpus analysis also makes it possible to compare the distribution of structural patterns across lexical categories, speakers, and grammatical environments.

Naturalistic data also contribute to psycholinguistic research by providing a necessary counterpoint to experimentally induced language switching. Blanco-Elorrieta and Pylkkänen (2017) demonstrated that the cognitive demands associated with voluntary bilingual language switching differ from those observed under externally cued switching conditions. Their later discussion of ecological validity in bilingualism research further argues that conclusions drawn from highly controlled experimental paradigms should not automatically be generalized to bilingual behaviour in natural interaction (Blanco-Elorrieta & Pylkkänen, 2018). This is also consistent with the Adaptive Control Hypothesis by Green and Abutalebi (2013) and the work of Beatty-Martínez et al. (2020), who similarly demonstrate that the consequences of bilingual experience depend heavily on interactional context, reinforcing the methodological value of studying bilingual speech in environments where language alternation arises from speakers' own communicative behaviour.

Conversational approaches to code-switching have long emphasized this contextual dimension. For example, Auer (1998) treats switching as an interactionally meaningful phenomenon whose interpretation depends on its position within the conversation, while Gardner-Chloros (2009) highlights the complex relationship between linguistic structure, speaker characteristics, and social context. Although the dissertation is mainly concerned with grammatical rather than conversational functions, such works are important to mention because they demonstrate that naturally occurring switches are embedded within communicative events.

2.6 Hungarian-English Language Contact and Code-Switching

Hungarian-English language contact has been investigated from several perspectives, including minority-language maintenance, immigrant bilingualism, lexical borrowing, code-switching, and morphosyntactic integration. Compared to language pairs such as Spanish-English or French-English, however, it remains a relatively less extensively documented phenomenon. Nonetheless, the existing literature is substantial enough to establish both a sociolinguistic background and a number of recurrent structural patterns that are relevant to the present study.

2.6.1 Sociolinguistic Background of Hungarian-English Contact

Research on Hungarian in contact with other languages has developed through the study of Hungarian-speaking minority and immigrant communities outside Hungary. Fenyvesi's (2005b) edited volume *Hungarian Language Contact Outside Hungary* represents one of the most comprehensive works examining Hungarian as a minority language in communities located both in neighbouring countries and overseas, and addresses language-contact effects at both sociolinguistic and structural levels. Her volume highlights the relationship between patterns of language use, community circumstances, and structural characteristics of Hungarian under sustained contact with a majority language.

This contact has been documented in immigrant communities in English-speaking countries. While Fenyvesi's (2005a) study of Hungarian in the United States examines Hungarian as a heritage and minority language in an English-dominant sociolinguistic environment, M. Kovács (2005) reports on Hungarian in Australia. These works offer evidence of sustained contact in which Hungarian speakers have regular access to English and display borrowing, code-switching, structural influence, and processes associated with language maintenance or shift.

Fenyvesi's (2005b) volume approaches this language contact at both macro-sociolinguistic and micro-linguistic levels and draws attention to the role of typological properties in determining possible outcomes of the contact. In the same volume de Groot (2005) considers the contact from a linguistic-typological perspective, reinforcing the importance of analysing the different structural outcomes in relation to the grammatical characteristics of Hungarian itself.

However, English influence on Hungarian is not limited to immigrant bilingualism. Anglicisms in Hungarian have also been documented by Farkas and Kniezsa (2004), who observed the incorporation of English lexical material into Hungarian used in Hungary. In their overview of English influence on Hungarian, they discuss the history of this language contact together with pronunciation, orthography, morphology, semantic change, and different forms of borrowing. They demonstrate that English-origin lexical material may undergo adaptation at several linguistic levels in Hungarian. More recent work by Vishogradska (2017) shows that English loanwords can display systematic morphophonological adaptation while at the same time retaining properties that mark them as relatively foreign within Hungarian.

The sociolinguistic conditions under which English material becomes available to Hungarian speakers may affect its frequency, conventionalization, distribution across speakers, and eventual status as borrowing, established loanword or code-switching.

2.6.2 Previous Research on Hungarian-English Code-Switching

A substantial part of specifically Hungarian-English code-switching literature has been based on Hungarian-American bilingual communities and has approached switching from sociolinguistic, pragmatic, grammatical, or cognitive perspectives. Myers-Scotton and Bolonyai (2001) analyse code-switching within the Rational Choice Model. They argue that linguistic choices can be understood as individually made but socially constrained decisions where speakers select among the available linguistic alternatives based on their communicative goals, costs and benefits. Combining Hungarian-English data with theoretically explicit analyses of bilingual grammar, Bolonyai's (2005) study *English Verbs in Hungarian/English Codeswitching*, investigates the grammatical integration of English verbs into Hungarian. Her focus is on why Hungarian employs morphological integration when incorporating English verbs.

Bhatt and Bolonyai (2011) extend this through the Optimality-Theory of bilingual language use. They argue that code-switching emerges from the interaction of ranked grammatical and socio-pragmatic constraints rather than from a single categorical rule. Their study shows that Hungarian-English data have already been used to connect formal structure with the communicative optimisation of bilingual choices. The same year, T. Kovács (2011) also contributed to the grammatical and socio-pragmatic analysis code-switching by evaluating the applicability of Optimality Theory to Hungarian-English grammar and examining whether code-switching patterns can be accounted for through interacting constraints. This work was later refined in T. Kovács's (2018) monograph, which combines linguistic-pragmatic and socio-cultural considerations and draws on a corpus of personal interviews of a Hungarian-American bilingual community in North Carolina.

Thus, Hungarian-English code-switching has been shown to be relevant to both sociolinguistic and formal questions, however, relatively few studies have subjected the same corpus of Hungarian-English naturalistic data to systematic comparison under competing morphosyntactic frameworks.

2.6.3 Morphological Integration of English Elements in Hungarian

As Hungarian, being an agglutinative language, its grammar often requires overt morphological realization. It has rich nominal and verbal morphology, including case suffixes, possessive marking, verbal inflection, derivational morphology, and preverbal elements. Bolonyai (2005) contrasts its verbal morphology with English in explaining the integration of English verbal material. The grammatical outcomes of this contact are often clearly visible on individual mixed words created by Hungarian verbalizing morphology. Using the data of Kontra (1995) and others, she examines English verbal stems being followed by a Hungarian verbalizer, after which the necessary inflectional morphology can be attached. Table 2 presents such examples (Bolonyai, 2005, p. 320), demonstrating that English-origin lexical bases can participate in a fully Hungarian inflectional structure.

Table 2. Examples of English verbal material integrated into Hungarian from Kontra (1995)

Ex.4	*'És utána te* decide*-ol-hat-sz'* (tr.: *And then you can decide*)
Gl.:	and then you decide.V-VBZ-POT-PR.2S
Ex.5	*'Ki-*rent*-ol-ja nek-em a bézment-et'* (tr.: *She rents the basement to me*)
Gl.:	out.COV-rent.V-VBZ-PR/3S/DEF DAT-POSS.1SG the basement-ACC

In the examples, Bolonyai (2005) relates this pattern to the ordinary derivational resources of Hungarian, which uses the verbalizers *-l* and *-z*, together with their phonologically conditioned variants in deriving verbs. This bilingual verbalizer is not an arbitrary device introduced for mixed speech, but an already available morphological mechanism in Hungarian grammar which overtly establishes verbal categorial identity and satisfies the structural requirements of Hungarian. So, the outcome is not merely the juxtaposition of an English word and a Hungarian suffix but a construction where category assignment, derivation, and subsequent inflection interact.

Earlier explanations of the phenomenon were of different positions: Bolonyai (2005) reviews Moravcsik's (1975) proposal that other-language verbal material enters Hungarian through nominal status and because of this, it requires a verbalizer. On the other hand, she argues that the verbalizer has a more substantial grammatical role, because it overtly marks the derived

form as a verb and thereby fulfils the requirement for Hungarian verb formation. Also, she interprets this verbalizer within Myers-Scotton's framework and treats Hungarian as the ML and English as the EL in these constructions.

The relevance of such data, however, extends beyond verbal material. The work of Farkas & Kniezsa (2004) on English borrowing into Hungarian shows that English-origin material may undergo both morphological and phonological adaptation. In morphologically rich languages, nouns and adjectives may participate in recipient-language grammar, taking, for example, case, number, possessive, derivational, or degree morphology if need. Although the different grammatical categories should not be assumed to behave the same way.

Differentiating and classifying cases of borrowings and code-switches has been attempted by Bolonyai (2005) by referring to morphosyntactically integrated English forms that retain English-like pronunciation as CS, while forms displaying both morphosyntactic and phonological integration as borrowings for the purposes of her study. This suggests that even within the Hungarian-English literature, the same observable property, i.e. integration, does not necessarily point to the same classification of the English-origin material.

2.6.4 Research Gaps

Studies of Hungarian minority communities have documented the sociolinguistic consequences of sustained contact with majority languages and have demonstrated that Hungarian may undergo lexical and structural influence under bilingual conditions (Fenyvesi, 2005b). Research specifically concerning English influence has described the incorporation and adaptation of Anglicisms in Hungarian (Farkas & Kniezsa, 2004), while Hungarian-English code-switching studies have examined the sociolinguistic motivation, pragmatic organization, and grammatical properties of bilingual language use (Bolonyai, 2005; T. Kovács, 2011, 2018; Myers-Scotton & Bolonyai, 2001). Bolonyai' (2005) analysis also demonstrates how such patterns can be interpreted within the Matrix Language Frame tradition. What remains considerably less developed is a systematic comparison of how the same integration patterns are interpreted by competing theoretical approaches.

This gap in Hungarian-English analysis also proves the need for a distinction between applying a model and comparing models. Previous studies have successfully demonstrated that individual theoretical approaches can shed light on the different aspects of Hungarian-English bilingual speech; however, showing that a dataset is compatible with one framework does not

necessarily mean that the framework provides a better account than an alternative one. Therefore, a comparative analysis that requires the predictions of competing approaches which are then evaluated against the same set of observations becomes necessary.

Also, methodology-wise, much of the earlier Hungarian-English literature has dealt with Hungarian used by immigrant or heritage-language communities in English-dominant environments, while the present corpus, consists of naturally occurring, predominantly Hungarian discourse where English-origin material appears only occasionally. This provides a different contact setting and allows the examination of recurring integration patterns without eliciting the switches.

Another gap concerns the different grammatical categories of the English material and their morphosyntactic structures subjected to the comparison. As discussed, the integration of verbs has received a lot of attention, but examining additional domains (including nominal case marking, possession, plural morphology, adjectival derivation, degree marking, mixed compounds, and constructions involving verbal prefixes or coverbs) makes it possible to determine whether the same theoretical predictions perform consistently across categories or whether its success depends on the type of integration.

Addressing these gaps makes it possible to move beyond demonstrating that a given model can accommodate selected Hungarian-English examples and instead evaluate where the competing approaches converge, where they differ, and which aspects of Hungarian-English bilingual grammar remain difficult for either framework to explain.

3. MORPHOSYNTACTIC INTEGRATION

3.1 Descriptions of Morphosyntactic Integration

Morphosyntactic integration refers to the extent to which linguistic material originating in one language conforms to the morphological and syntactic requirements of another language when it is incorporated into bilingual discourse. In contact-linguistic research, integration is therefore not limited to the attachment of overt morphology (Deuchar & Stammers, 2016; Poplack, 2018). It may involve several related processes, including the assignment of grammatical category, the use of recipient- or Matrix-Language derivational morphemes, inflections, and the accommodation of the item to language-specific word-order requirements. Thus, other-language elements can be integrated at more than one grammatical level and to different degrees.

This phenomenon has played an important role in attempts to distinguish code-switching from borrowing. Earlier contact-linguistic work already recognized that lexical material from a language may undergo structural adaptation upon entering another one (Haugen, 1950; Weinreich, 1953). Later variationist research made the grammatical behaviour of other-language material the main empirical criterion. The works of Poplack and Sankoff (1984), Poplack et al. (1988), and Poplack (2018) emphasize that the relevant question is not simply whether an item carries an overt recipient-language affix, but whether it patterns morphosyntactically the same way as the comparable recipient-language lexical items. Distinguishing the two is important because recipient-language grammatical marking itself can be variable, and the absence of overt affixation does not necessarily imply the absence of integration if the native forms exhibit the same variation.

The same-time occurrence of both morphosyntactic- and phonological integration is possible but not necessary. An other-language lexical item may take on morphology while retaining the phonetic characteristics of its source language, or on the contrary, it may show considerable phonological adaptation without clear evidence concerning its grammatical status. Applied to the current language pair, experimental work by Hayes et al. (2009) found that Hungarian speakers possess productive and probabilistic knowledge of vowel-harmony patterns and can use it to phonologically integrate novel forms. Siptár and Törkenczy (2000) describe these alternations where the selection of appropriate suffixes on English-origin bases serves as evidence that speakers actively accommodate new lexical material to Hungarian

morphophonological rules, although as discussed earlier, it does not determine whether the item is classified as a borrowing or a code-switch.

The grammatical interpretation of morphosyntactic integration may vary according to theoretical framework. In the MLF Model, morphosyntactic asymmetry is a basic property of most mixed constituents where the Matrix Language supplies the grammar into which Embedded Language material can be incorporated under specified structural conditions (Myers-Scotton, 1993, 2002). On the other hand, the Poplackian approach requires morphosyntactic patterning to determine the borrowing status of the donor-language item. Also, the findings of Budzhak-Jones (1998) on her analysis of Ukrainian-English bilingual discourse illustrate that an intraword language boundary is an observable structural property, but its classification as code-switching cannot be assumed solely from the different lexical origins of the stem and the affix, so she applied the term borrowing to such cases.

Finally, the source of the lexical stem and the language responsible for its morphosyntactic behaviour does not need to coincide, as shown by Bolonyai (2005), who also argues that the verbalizer (*-l, -z*) functions as an overt marker of verbal categorial identity and relates this requirement to ordinary Hungarian word formation, as mentioned earlier. Consequently, integration should be treated as a multidimensional property rather than as a distinction between "integrated" and "unintegrated" forms. English-origin items may undergo Hungarian syntactic positioning without overt morphological integration, receive Hungarian inflection without extensive phonological adaptation, or a combination of morphological, syntactic, and phonological adaptation. The more intriguing area of research interest could be multiword English material that may preserve more of its internal English structure, while its individual stems can become embedded into Hungarian grammar, as apparent in the examples of Table 3.

Table 3. Examples of multiword English material integrated into Hungarian from the data

Ex.6	*'most a main game-ek-et nem néz-ik'* (tr.: '*now the main games aren't watched'*)
Gl.	now.ADV the.DET.DEF main.ADJ game.N-PL-ACC no.NEG watch.V-3PL.ACC
Ex.7	*'lehet kér-ni join request-et'* (tr.: '*you can ask for a join request'*)
Gl.	may.AUX ask.V-INF join.V request.N-ACC

The term *morphosyntactic integration* does not predetermine whether the resulting form is classified as a code-switch or a borrowing. This classification depends on the theoretical criteria considered in Chapter 4; the purpose here is to establish the Hungarian grammatical structures in which the integration of English-origin material can be examined.

3.2 Morphosyntactic Integration in Comparable Contact Settings

As discussed earlier, contact involving morphologically rich recipient or matrix languages provides especially visible evidence of how donor-language material is incorporated into another grammatical system. Case marking, possession, category-forming suffixes, verbal inflection, morphophonologically ruled allomorphy, and other bilingual contact phenomena create points at which the contribution of each language can be examined.

Bilingual Finnish, Turkish, and Welsh data provide the most relevant comparison for the dissertation. Finnish offers one of the closest typological comparisons with Hungarian, while Turkish gives evidence from another, less related agglutinative language. Welsh-English research, however, has played a central role in the theoretical debate concerning code-switching, borrowing, and grammatical integration. Estonian and Croatian data must be considered as well, since they offer additional evidence that integration may work differently across linguistic levels.

3.2.1 Finnish-English Contact

Finnish-English research provides a particularly useful comparison because Finnish, like Hungarian, has an extensive case system and productive suffixal morphology. Poplack et al. (1989) used Finnish-English data to distinguish code-switching from borrowing within a variationist framework. They argue that LOLIs in Finnish should be treated as borrowings, as they followed Finnish grammatical patterns.

Halmari's (1997) study of American Finnish likewise demonstrated that English-origin nouns frequently receive Finnish case morphology according to the syntactic requirements of the containing clause. She uses a formal syntactic framework, more specifically a modified version of the Government Constraint, to explain that head-complement relationships play a primary role in predicting where code-switching can easily occur. Her findings suggest that bilingual speech is constrained by structural grammar, proving that syntax cannot be left out when analysing contact phenomena.

3.2.2 Turkish Contact Morphology

Agglutinating languages outside the Uralic language family show evidence of the same linguistic behaviour, despite the typological differences. Like Hungarian, Turkish bilingual data similarly demonstrate extensive morphosyntactic integration. Backus (1992, 1996) documented the insertion of Dutch lexical material into Turkish grammar, including the integration of nouns and verbs.

While Hungarian grammar applies its verbalizing suffixes, Turkish predicates may be combined with the native light verb *yapmak* 'do/make' (Backus, 2009). A similar mechanism has been observed in Spanish-English bilingual speech as well, using the verb *hacer* 'do/make' combined with bare English verbs (Wilson & Dumont, 2015). In such cases, other-language lexical material must receive a category compatible with the grammatical system of the language into which they are integrated.

The description of the Turkish-Dutch language pair by Demirçay and Backus (2014) and Backus and Demirçay (2021) distinguishes between lone and longer multiword insertions and suggests that constituent size and grammatical integration should be treated as separate dimensions, because the more complex or multiword Dutch material has also been integrated into the Turkish structure.

3.2.3 Welsh-English Contact

Research on Welsh-English has been used to examine grammatical congruence, the MLF Model, and the Nonce Borrowing Hypothesis as well. Deuchar (2005) showed that the integration of English-origin material into Welsh depends partly on the degree of grammatical congruence between the two languages. The availability of a corresponding category or structural position in Welsh influences how readily an English-origin noun, adjective, or verb can be incorporated.

Later, Deuchar (2006) applied the MLF Model to Welsh-English conversation data and found strong support for the identification of a Matrix Language and for an asymmetrical division between the grammatical and lexical contributions of the languages, providing an important comparison with language pairs in which a morphologically rich language supplies the grammatical frame.

Welsh-English data have also been central to a debate surrounding the Nonce Borrowing Hypothesis. Stammers and Deuchar (2012) examined the application of Welsh soft mutation to English-origin verbs and found that the degree of integration was related to lexical frequency. When Poplack (2012) questioned whether this represented an appropriate test of the NBH, Deuchar and Stammers (2012) defended the relevance of frequency to the empirical evaluation of the hypothesis.

Debates of such nature are crucial for testing the validity of linguistic theories. Applying their predictions and principles can demonstrate that integration is not always interpreted the same way. Evidence from morphosyntactic integration may be used to support an account based on recipient-language borrowing, Matrix-Language incorporation, or gradient lexical establishment.

Deuchar and Stammers (2016) also compared the approaches of Poplack and Myers-Scotton on English-origin verbs in Welsh. They show that the apparent degree of integration depends on whether morphological, syntactic, morphophonological, frequency, or dictionary-based diagnostics are applied. Precedents like theirs are especially important for the dissertation because they demonstrate that frameworks may agree on observable structure while they differ in the significance assigned to them.

3.2.4 Additional Evidence from Estonian and Croatian

Examining the Russian-Estonian language pair, Zabrodskaja (2009) found multiple degrees of morphological integration instead of an opposition between fully integrated and completely unintegrated items. While asserting that integration is not a binary phenomenon, being present or absent, Zabrodskaja and Verschik (2014) gave further evidence that morphological and phonological adaptation do not necessarily develop together.

The morphosyntactic integration of English material has been documented in the Croatian-English contact too. Hlavac (1999, 2003) documented variable phonological and morphosyntactic adaptation of English-origin material in Croatian bilingual speech. In Hlavac's (1999) earlier study, he characterizes phonological integration as a continuum involving partial and complete accommodation, as sounds such as /θ/, /ð/, /w/, and /ŋ/ were especially resistant to transphonemisation. His examples from 2003, however, show cases of clear phonological and morphological integration, observable in the example *kontejner-e* 'containers' (Hlavac, 2003, p. 207), although the two do not necessarily come together, as seen in city-ju 'city-ACC'

(p. 202), where recipient-language morphology can be attached even when the lexical base retains its English phonology. These forms thus highlight the multidimensional characteristics of linguistic integration.

3.3 Structural Properties of Hungarian Relevant to Code-Switching

Since Hungarian is commonly characterized as a morphologically rich, predominantly agglutinative language, with extensive affixation in both nominal and verbal paradigms (É. Kiss, 2002; Kenesei et al., 1998; Kiefer, 2000), it is particularly suitable for investigating morphosyntactic integration; even in literary works, where such integration might carry additional functions and communicative goals, resulting in further classification (Domokos & Deganutti, 2021; T. Kovács, 2024).

Typological distance between English and Hungarian has direct consequences for bilingual constructions: an English lexical item inserted into Hungarian discourse can enter environments where Hungarian requires overt realization of grammatical properties that English either expresses analytically or does not mark on the corresponding item.

3.3.1 Word Order

As the word order of Hungarian is much more flexible than that of English, its syntax has discourse-configurational organization; constituent position is closely related to information-structural functions such as topic and focus, and it is not determined exclusively by grammatical relations. É. Kiss's (1987, 1995, 2002) analyses of Hungarian syntax place the topic – predicate articulation of the sentence and the preverbal focus field at the centre of Hungarian constituent structure.

Focused constituents occupy a structurally privileged position immediately preceding the verb as É. Kiss (1998, 2002) reports. She also observes that the position of verbal prefixes or coverbs is also sensitive to these information-structural configurations, since a coverb typically immediately precedes the verb in a neutral clause; however, focus, negation and certain other operators can relocate it into a postverbal position. Surányi's (2009) detailed analysis of Hungarian verbal constituents further distinguishes their position inside the verbal phrase from configurations where they occur in the preverbal functional field, which emphasises that particle placement is structurally conditioned.

Additionally, Demszky (2021) demonstrates that Hungarian verb – object order is affected not only by information-structural variables but also by object definiteness, the weight of the noun phrase, and the lexical-semantic class of the verb. This means that Hungarian word order reflects the interaction of syntax, information structure, and predicate organization. Despite being relatively flexible, it is far from allowing totally unrestricted variations.

This flexibility is relevant to code-switching because the position of English-origin material cannot be evaluated merely by comparing its linear order with English SVO syntax. An English-origin constituent may occupy a position determined by Hungarian topic – focus organization and may interact with preverbal elements and operators, although word order alone is unlikely to provide an unambiguous classification of (non-)integrated forms.

3.3.2 Nominal Morphology

Hungarian nominal morphology provides quite transparent evidence for structural integration because nouns may carry plural, possessive, and case morphology, and several morphemes can occur within the same word (Kenesei et al., 1998; Kiefer, 2000), so English-origin lexemes can be integrated and receive Hungarian nominal inflections.

Case marking offers especially clear evidence. Hungarian case suffixes overtly encode grammatical, spatial, and directional relations, and many exhibit vowel-harmonic allomorphy. In mixed forms, the English-origin stem supplies the lexical material while the Hungarian suffix realizes its grammatical relation within the clause. In Bolonyai's analysis (2005, p. 320), *bike-jack-et*, is an example where the English nominal base receives the Hungarian accusative suffix *-t*. Many comparable structures occur in the present corpus with English-origin stems carrying Hungarian case and possessive morphology (See Table 4).

Table 4. Examples of Hungarian nominal morphology applied on English stems

Ex.8	*'jam session-ök-be'* (tr.: *into jam sessions*)
Gl.:	jam.N session.N-PL-ILLAT
Ex.9	*'fight-ok-at'*
Gl.:	fight.N-PL-ACC
Ex.10	*'challenge-ünk'*

Gl.:	challenge.N-POS.1PL

As presented above, Hungarian plural, possessive, and case markers can be combined in ordered morphological sequences, which allows mixed forms to contain more than a single layer of Hungarian grammatical material. Such structures are informative because the contribution of each language can be identified at the morpheme level already, without the need for constituent-level syntax analysis. Recent work on mixed nominal structures (Alexiadou, 2024) likewise demonstrates the theoretical value of number morphology for analysing how grammatical material from different languages is put together within a single nominal expression; still, the presence of morphological integration alone does not determine the status of such forms as a borrowing or a code-switch.

3.3.3 Verbal Morphology and Verbalization

The verbal domain is also quite revealing in Hungarian-English contact because Hungarian requires substantially more overt verbal morphology. Hungarian verbs inflect for person and number, tense and mood, while the finite conjugational paradigm also interacts with the definiteness of the object; verbal prefixes and derivational morphology additionally contribute to aspectual and lexical meaning (Bolonyai, 2005). English-origin verbal material integrated into this system must accommodate grammatical distinctions that are otherwise not morphologically expressed on the corresponding English verb.

Table 5. Examples of English bases integrated by Hungarian verbalizers

Ex.11	*'coach-ol-sz'*
Gl.:	coach-VBZ-2S
Ex.12	*'hype-ol-t'*
Gl.:	hype-VBZ-PAST

A particularly important mechanism is verbalization. Hungarian productively derives verbs through the suffixes *-l* and *-z* and their vowel-harmonic variants, including *-ol/-el/-öl* and *-oz/-ez/-öz*. Unlike English, where noun – verb conversion can occur without overt morphology, Hungarian marks the formation of new denominal verbs morphologically (Bolonyai, 2005). In

Hungarian-English bilingual speech, this mechanism extends to English verbal bases, producing the characteristic structure:

English lexical base + Hungarian verbalizer + Hungarian inflection(s)

Bolonyai (2005) documents forms having this structure (see Tables 2 & 5) and analyses the verbalizer as an overt marker of verbal categorial identity: the English base is first incorporated through derivation, after which Hungarian inflection (and if needed, coverbs) can be added. This corresponds to Wohlgemuth's (2009) indirect insertion strategy, in which other-language verbal bases receive overt verbalizing elements before additional recipient-language inflections. However, the morphosyntactic integration of verbs may remain sensitive to factors such as finiteness and morphological markedness as shown by Shaw & De Smet (2022).

3.3.4 Adjectival Morphology and Recategorization

Adjectival morphology provides another domain in which category assignment can become overt. Hungarian productively derives adjectives through suffixation, particularly with *-s* and forms containing a linking vowel such as *-os, -as, -es,* and *-ös* (Kenesei et al., 1998; Kiefer, 2000). When such morphology attaches to an English-origin base, it does not only integrate an existing English adjective: it establishes adjectival status within Hungarian.

Hungarian adjectives can also participate in degree morphology (see Table 6). Comparative forms are marked by the suffix *-bb* and its phonologically conditioned forms, while superlatives combine the prefix *leg-* with the comparative. Consequently, an English-origin base can undergo Hungarian adjectival derivation and degree formation, producing a multilayered structure in which recategorization precedes receiving additional Hungarian morphology. Such forms make it necessary to distinguish the lexical category of the English source item from its grammatical category after integration.

Table 6. Examples of English-origin items receiving adjectival morphology

Ex.13	*'leg-light-os-abb-ak'*
Gl.:	SUP-light-ADJZ-COMP-PL
Ex.14	*'heavy-bb'*
Gl.:	heavy-COMP

This distinction is supported more broadly by Meechan and Poplack's (1995) analysis of Wolof-French and Fongbe-French bilingual speech, since their comparison demonstrates that categorial mismatches between the languages constrain mixing and requires language-specific strategies for establishing categorial compatibility. These English-origin bases combined with Hungarian adjectival morphology are theoretically informative because they make category assignment and morphological integration clearly visible.

3.3.5 Coverbs / Verbal Prefixes

Hungarian verbal prefixes (or coverbs), such as *meg-*, *el-*, *ki-*, *be-*, *fel-* and *le-*, contribute directional, aspectual, and lexical-semantic information to the verb. Unlike inseparable prefixes, however, their position is syntactically variable: they typically occur immediately before the verb in neutral clauses but may be separated from it in constructions involving focus, negation, and other preverbal operators (Surányi, 2009; Kalivoda, 2022). Their distribution is relevant not only to meaning but also to Hungarian clause structure and information structure.

This becomes particularly informative in bilingual predicates. Bolonyai (2005) describes constructions such as *el-miss-ol-t-am* 'away.COV-miss-VBZ-PAST-1S' and *ki-rent-ol-ja* 'out.COV-rent-VBZ-1S.DEF', in which a Hungarian coverb combines with an English lexical base, a Hungarian verbalizer, and Hungarian inflection. The structure of these forms can be represented as follows:

Hungarian coverb + English lexical base + Hungarian verbalizer + Hungarian inflection

Their significance extends beyond word-internal morphological integration. Since these Hungarian coverbs participate in independent syntactic and aspectual structures, their combination and positioning with English-origin verbal material can indicate that the mixed predicate is being organized according to Hungarian predicate and clause-level grammar (Surányi, 2009; Kardos & Farkas, 2022). Thus, they serve as a bridge between morphological and syntactic integration, without requiring commitment to any concrete CS model.

3.3.6 Determiners and Nominal Constructions

Hungarian definite and indefinite articles (*a/az* and *egy*), together with quantifiers, modifiers, possessive elements, and case morphology, occur in systematically organized nominal structures (É. Kiss, 2002). An English-origin noun occurring with a Hungarian determiner (frequently with additional Hungarian case or possessive morphology) can occupy a position defined by Hungarian nominal grammar, as seen in the examples of Table 7 below.

Table 7. Examples of integrated English items with Hungarian determiners

Ex.15	*'lát-tá-tok a progress-t'* (tr.: you saw the progress)
Gl.:	see.V-PAST-3PL.ACC the.DET.DEF progress.N-ACC
Ex.16	*'a french fries-ra'*
Gl.:	the.DET.DEF french.ADJ fries.N-SUBLAT

An isolated English lexical head embedded in an otherwise Hungarian nominal structure must be distinguished from a multiword constituent that preserves English-internal organization. Mixed noun phrases are a well-established conflict domain in code-switching research, with determiner language, adjective – noun order, and the distribution of grammatical morphology providing evidence on how the two grammatical systems are combined (Parafita Couto & Gullberg, 2019). For this study, the relevant diagnostics include determiner selection, internal constituent order, and the locus of Hungarian case marking.

Nominal integration may have consequences beyond the noun phrase, as Hungarian transitive verbs alternate between indefinite and definite conjugation according to grammatical properties of the direct object, although the distribution cannot be reduced to definiteness alone (Coppock, 2013). Consequently, an English-origin object may participate in a dependency where its grammatical interpretation is reflected in Hungarian verbal morphology, which is informative because the integration of the English-origin constituent is also observable through the grammatical relations across the clause.

3.3.7 Compounding and Mixed Compounds

Compounding is a further domain where English-origin material can participate in Hungarian word formation. Hungarian has a productive compounding system, and similarly to English, endocentric compounds are right-headed, meaning that the rightmost constituent determines the syntactic category of the compound (Kiefer, 2000, 2012). Examples of Hungarian-English compounds are illustrated in Table 8.

Table 8. Hungarian-English compounds from the data

Ex.17	*'review videó-k-at'*
Gl.:	review video.N-PL-ACC
Ex.18	*'normális content gyártás-sal'*
Gl.:	normal.ADJ content production.N-INSTR
Ex.19	*'scam levél'*
Gl.:	scam letter.N

It is important to distinguish mixed compounds formed within a Hungarian compound structure from internally English compounds or larger lexical units inserted into Hungarian discourse. Cross-linguistic research supports this distinction. For example, Treffers-Daller (2005) shows that French-Dutch mixed compounds in Brussels Dutch conform to Dutch compound structure, whereas inserted French nominal groups preserve their internal French organization. Alexiadou (2020) also distinguishes mixed compounds constructed according to the grammar of one language from compounds inserted as already formed units. More recent experimental evidence from Pierantozzi & Cocchi (2024) confirms that mixed compounding is possible but may be constrained by incompatibilities between the language pair (e.g., differences in head position or inflectional properties).

For the present dissertation, mixed compounds are useful, because they reveal how English-origin material participates in Hungarian word formation and their analysis can provide evidence for the classification within the compared frameworks as code-switches or borrowings.

3.4 Structural Configurations of English-Origin Material

English-origin material in Hungarian discourse occurs in structurally different configurations, ranging from lone lexical items and intraword mixed forms to mixed constituents, mixed compounds, and internally English multiword sequences. These configurations are treated here as descriptive categories rather than predetermined contact types. This distinction is necessary because constituent size and structural integration on their own do not categorise bilingual forms as borrowings or code-switches. Muysken (2000), for example, distinguishes insertion from larger-scale alternation, while the MLF and variationist traditions also distinguish lone material from internally structured multiword constituents, although they interpret and identify these forms differently (Myers-Scotton, 1993, 2002; Poplack & Meechan, 1998; Poplack, 2018).

Lone items may occur with or without overt Hungarian morphology, and their grammatical behaviour is much more relevant to classification (Poplack, 2018) than their occasional isolated occurrence. Intraword forms make the interaction of the languages particularly visible, as English-origin bases may receive Hungarian derivational and/or inflectional morphology, illustrated by the examples of Tables 4, 5 and 6. Morphological mixing does not determine contact status independently since comparable forms have been analysed as borrowings in some approaches (Budzhak-Jones, 1998) and as code-switches in others.

As discussed earlier, longer other-language structures require attention to their internal grammatical organization. Multiword material is not necessarily phrasal code-switching simply because it contains several orthographic words, since compounds, collocations, and lexicalized sequences may function as complex lexical units (Backus, 2003). In the MLF Model, English multiword constituents that preserve English-internal dependencies may correspond to an Embedded Language island (Jake & Myers-Scotton, 1997; Myers-Scotton, 2002), whereas mixed compounds constitute another distinct configuration because material from both languages participates in a word-formation relationship. According to Treffers-Daller (2005), internally complex material must be distinguished from both lone insertion and phrasal alternation.

Constituent size, internal grammatical organization, external Hungarian integration, and morphological complexity are considered as somewhat independent dimensions. Isolated items may be extensively integrated morphologically, while multiword English expressions may

retain their internal structure and receive Hungarian marking. Therefore, the dissertation distinguishes among LOLIs, intraword mixed forms, mixed constituents, mixed compounds, and internally English multiword constituents. These descriptive categories provide a common basis for the framework-specific analyses in Chapter 4 without assuming in advance whether any individual construction is a borrowing, code-switching, insertion, alternation, or an Embedded Language island.

4. COMPETING ACCOUNTS OF BILINGUAL MORPHOSYNTAX

4.1 Rationale for the Comparison of Theoretical Frameworks

The grammatical study of bilingual speech has produced several theoretical approaches to the structural organization and classification of mixed-language material. The present dissertation compares two particularly influential traditions: the Matrix Language Frame (MLF) Model and the variationist approach associated with Poplack. The originality of the study lies in the fact that both are applied to the same empirical questions: How is morphosyntactically integrated English-origin material structured and analysed in Hungarian-English bilingual speech, and how successfully does each framework account for the observed constructions?

Comparing the two is important because they assign different theoretical significance to the same structural evidence. As already discussed, the MLF Model analyses bilingual constructions in terms of structural asymmetry, proposing that a Matrix Language provides the morphosyntactic frame where Embedded Language material can occur (Myers-Scotton, 1993, 2002), while the Poplackian approach evaluates donor-language material primarily through its grammatical behaviour in the recipient language, particularly when distinguishing code-switching from established or nonce borrowing (Poplack & Sankoff, 1984; Poplack et al., 1988; Sankoff et al., 1990; Poplack, 2012, 2018). Therefore, an English-origin element carrying Hungarian morphology constitutes the same empirical observation for both frameworks, but their predictions, diagnostics, and analytical status assigned to the structure differ.

Again, Hungarian-English data provide a particularly suitable basis for this comparison because Hungarian overtly realizes grammatical information through most morphosyntactic structures. During their evaluations neither approach is treated as theoretically superior in advance. A direct comparison is required by establishing a shared descriptive basis and independently formulated predictions. Although this method is far from being novel, applied to Hungarian-English corpora it has not been systematically studied and published yet according to the knowledge of the author. Deuchar and Stammers (2016) demonstrate the value of theory-neutral initial coding in their comparison of the Poplackian and Myers-Scotton approaches, while Duran Eppler et al. (2017) similarly show how competing frameworks can be operationalized and tested on the same corpus.

Therefore, this study first establishes the observable structural properties of the Hungarian-English constructions and then applies the predictions of each framework separately. The goal is not only to determine whether a model can describe individual examples, but to compare how consistently and precisely the two approaches predict and analyse the same patterns of morphosyntactic integration, where their analyses converge or diverge, and which constructions remain problematic or indeterminate.

4.2 The Matrix Language Frame Model

The Matrix Language Frame (MLF) Model is one of the most influential accounts of the morphosyntactic organization of intrasentential code-switching. Developed and later refined by Myers-Scotton (1993, 2002), the model proposes asymmetrical contribution to bilingual grammatical structure: one language provides morphosyntactic frame called the Matrix Language, while material from another is embedded.

Since English-origin lexical material frequently occurs with Hungarian derivational and inflectional morphology and within Hungarian syntactic structure, the framework is particularly relevant to the present data. The following sections outline the components of the model that generate testable predictions for such bilingual constructions.

4.2.1 Development and Scope of the MLF Model

The original MLF Model was developed to explain recurrent structural asymmetries in intrasentential code-switching and not necessarily to specify possible switch points (Myers-Scotton, 1993). The distinction between the ML and the EL was formulated primarily through the Morpheme Order Principle and the System Morpheme Principle. This was later expanded by the 4-M Model, discussed in detail in section 4.2.5.

Besides the 4-M Model, which provides a more detailed classification of morphemes, the Abstract Level Model, concerning different levels of lexical structure, the Uniform Structure Principle and the Asymmetry Principle are further developments of the MLF and should not be treated as components of a single unchanged model (Myers-Scotton & Jake, 2000; Myers-Scotton, 2002).

The MLF Model is intended primarily to account for what Myers-Scotton (2002) refers to as classic code-switching: bilingual complementizer phrases [CPs] in which one language provides an identifiable morphosyntactic frame while EL material occurs in it. For the current

study, a relevant question is whether the Hungarian-English constructions satisfy the predictions of classic code-switching, particularly whether Hungarian consistently supplies the grammatical frame.

4.2.2 The Matrix Language and Embedded Language

The ML – EL distinction expresses the central asymmetry of the model. Most morphosyntactic procedures governing the bilingual constituent are of the ML, whereas the EL has a more restricted role, which is typically to contribute by content morphemes or internally EL-structured constituents (Myers-Scotton, 1993, 2002).

ML status is a structural property, so it should not be equated automatically with a speaker's first, dominant, or socially majority language. Although those factors may correlate with ML selection, its identification concerns the grammatical organization of the bilingual CP. To do so, analysis of the morpheme order and the distribution of system morphology must be done (Myers-Scotton, 2002). Deuchar's (2006) description of Welsh-English, for example, operationalizes ML identification through the structural properties of bilingual clauses.

This distinction is directly applicable to English-origin items occurring with Hungarian morphology. If Hungarian independently satisfies the criteria for ML status, English lexical material can be analysed as EL material in a Hungarian frame. However, Hungarian cannot be identified as the ML solely because the speakers predominantly use Hungarian in the data.

4.2.3 The Morpheme Order Principle

The Morpheme Order Principle (MOP) predicts that where ML and EL morphemes occur together within a bilingual constituent, their surface order follows that of the Matrix Language (Myers-Scotton, 1993). This means that when EL lexical material occurs within a mixed constituent, it should not create an ordering pattern incompatible with the ML frame.

The strength of the MOP depends on structural contrast between the languages. A bilingual construction is much more informative when Hungarian and English permit or require different ordering in the given environment, and the mixed structure follows Hungarian. Where both languages independently require or allow the same order, the construction remains compatible with the MOP but provides weaker evidence for identifying the ML. It also should be noted that the MOP applies to ML+EL constituents but not to Embedded Language islands with an internal structure is of the EL. The importance of such contrast is demonstrated by Nguyen (2025) in

Vietnamese-English code-switching, where the extensive overlap in SVO ordering substantially reduces the ability of the MOP to identify the Matrix Language. Cases in which Hungarian and English permit the same order are therefore treated in the present study as compatible with, but not a diagnostic of the MOP.

4.2.4 The System Morpheme Principle

The second major component of the original Matrix Language Framework is the System Morpheme Principle (SMP). In its early formulation, it restricts the occurrence of EL system morphemes within ML+EL constituents and predicts that structurally relevant system morphology comes from the Matrix Language (Myers-Scotton, 1993).

However, the original distinction between content and system morphemes, proved too broad to capture the different grammatical functions and activation properties of system morphology. Therefore, the 4-M Model refined the category and made the SMP apply specifically to outsider late system morphemes instead of all grammatical morphemes (Myers-Scotton & Jake, 2000; Myers-Scotton, 2002), and this refinement is crucial for Hungarian, as the dataset includes constructions where English system morphemes are visible (see Table 9).

Table 9. Bilingual examples with English system-morphology from the data

Ex.20	*'french frie-**s**-nak'*
Gl.:	french.ADJ fry.N-s-PL(Eng)-DAT(Hun)
Ex.21	*'és a men's rights activism'*
Gl.:	and.CONJ the.DET.DEF men.N-POS(Eng) right.N-PL(Eng) activism.N

4.2.5 The 4-M Model

The 4-M Model distinguishes morphemes according to both their grammatical role and the way they are activated during language production (Myers-Scotton & Jake, 2000; Myers-Scotton, 2002). It categorises morphemes into four types: content morphemes, early system morphemes, bridge late system morphemes, and outsider late system morphemes.

Content morphemes are directly involved in expressing conceptual content and can include prototypical nouns, verbs, and adjectives. Early system morphemes are activated together with

their content-morpheme heads, and they strongly depend on information available at the lemma level. According to the 4-M, neither category is restricted to the Matrix Language, whereas late system morphemes are structurally assigned during grammatical formulation. Bridge late system morphemes express relationships internal to their constituent and, coming from the name, outsider late system morphemes depend on grammatical information located outside the constituent in which they are realized, so subject-verb agreement morphology represents a typical case of the latter.

The distinction is important because the later SMP with the 4-M Model predicts strong ML control over outsider late system morphemes, while allows early system morphemes and bridges to come from the EL. This means that the MLF model does not claim that all grammatical morphology should come from the Matrix Language. Applied to Hungarian, each relevant morpheme must be classified according to the grammatical information determining its realization.

It is important to mention that later work by Myers-Scotton and Jake (2014) further develops the relationship between lexical selection and grammatical realization. They argue that non-finite EL verbs can be incorporated more readily into ML frames because they impose fewer incompatible grammatical requirements than finite EL verbs, while ML morphology supplies the necessary inflections.

4.2.6 Embedded Language Islands

Longer stretches of EL material are not uncommon in bilingual data, however, the MLF Model permits constituents whose internal grammar is supplied entirely by the EL. An Embedded Language island refers to EL morphemes forming a constituent that is well formed according to EL grammatical requirements while being in a larger ML-framed structure (Myers-Scotton, 1993, 2002).

As EL islands differ from ordinary ML+EL constituents, the constraints of MOP and SMP are not required to be applied. The internal order or morphology of an EL island does not need to follow ML grammar.

For the present data EL-island status must be supported by evidence of internal English grammatical organization, as the mere adjacency of multiple English-origin words would be insufficient: compounds, titles or fixed expressions, may contain English elements without displaying productive English syntactic dependencies. An EL island example from the data can

be *egy ilyen running gag* ‘a.DET.INDEF such.ADJ running gag’ with a clear internal English modifier – head structure.

4.2.7 Congruence and the Blocking Hypothesis

Direct incorporation of EL material into an ML+EL constituent also depends on congruence between the lexical and grammatical specifications of the participating languages. Myers-Scotton (1993) treats congruence as compatibility across properties including grammatical category, semantic-pragmatic features, and argument structure.

In case of a lack of congruence, the ML might block the insertion of EL. This is called the Blocking Hypothesis. It predicts that direct insertion of the EL element may be prevented, however, a possible outcome is an EL island, which allows the EL item to occur together with grammatical material compatible with its own lexical requirements. The framework also recognises additional accommodation or compromise strategies, where incompatibilities are resolved through overt grammatical material (Myers-Scotton, 2002).

4.2.8 Critical Considerations of the MLF Model

Although the MLF Model provides testable predictions for bilingual morphosyntax, several aspects of its application require caution. One methodological issue concerns the identification of the Matrix Language. If the same morphosyntactic properties are used both to establish ML status and then to confirm Matrix Language control, the analysis might risk circularity. For this reason, ML identification and the testing of individual MLF predictions should therefore be kept analytically independent if possible.

A fairly recent empirical challenge is provided by Nguyen (2025), who tests the MOP and SMP on natural Vietnamese-English code-switching. As Vietnamese and English are both predominantly SVO with relatively few overt morphological conflict sites, the two principles often fail to distinguish which language is the Matrix Language. In her study substantial part of the corpus remains unaccounted for by clear ML identification. She also identifies problems in using reconstructed monolingual varieties as the baseline for mixed speech, in assigning problematic structures to a Composite Matrix Language, and in inferring the language membership of unexpressed elements. Her findings, however, do not show that the MLF Model is generally ineffective; they simply demonstrate that its diagnostic power can depend on the structural contrasts between the members of a particular language pair.

Another difficulty can be the empirical application of the 4-M Model. Its refinement of the original content – system distinction makes the System Morpheme Principle more specific, but it also requires meticulous classification of individual grammatical morphemes according to their dependencies, since labels such as case marker or determiner do not determine their 4-M status by themselves. As a result, difficult morpheme classification can affect the strength of conclusions drawn from the SMP (together with the 4-M).

As introduced in earlier chapters, the necessity of a distinct Matrix Language mechanism has been questioned by MacSwan (1999, 2000, 2005), who argues that bilingual constructions can be derived through independently motivated grammatical principles without code-switching-specific constraints. On the other hand, Jake et al. (2002) maintain that the structural asymmetries observed in bilingual clauses require explicit recognition of a Matrix Language. This shows that whether the asymmetries shown by the MLF Model constitute evidence for a specifically bilingual grammatical mechanism or can be derived from general properties of grammar is debated.

Still, these considerations do not prevent the MLF Model from being tested on the present data, but they affect how supporting evidence should be interpreted. In the analysis, conformity with Matrix Language ordering, the 4-M-informed System Morpheme Principle, or independently motivated EL-island structure will be treated as evidence of successful compatibility with the model.

4.3 The Poplackian Variationist Approach

Poplack's approach combines grammatical analysis with variationist methodology to distinguish different forms of bilingual language mixing empirically. Its early development focused on grammatical constraints on intrasentential code-switching, namely the Equivalence and Free Morpheme Constraints (Poplack, 1980; Sankoff & Poplack, 1981). Her and her colleagues' later work increasingly focused on the distinction between code-switching and lexical borrowing, especially the status of LOLIs occurring within recipient-language discourse (Poplack & Sankoff, 1984; Poplack et al., 1988, 1989; Sankoff et al., 1990; Poplack, 2012, 2018).

Although the Poplackian view approaches these structures through different diagnostics, the same empirical problem is examined: how English-origin material behaves when incorporated into Hungarian morphosyntax. This approach, however, places particular emphasis on whether

donor-language-origin material patterns grammatically with the recipient language and whether its distribution is characteristic of borrowing or code-switching.

4.3.1 The Equivalence Constraint

The Equivalence Constraint proposes that intrasentential switches tend to occur at points where the surface structures of the participating languages are compatible, so that juxtaposing material from the two languages does not violate the linear organisation of either grammar (Poplack, 1980; Sankoff & Poplack, 1981).

It differs from an account in which one language alone supplies the grammatical frame, as it evaluates the ordering requirements of both languages around the switching boundary. Applied for Hungarian-English, caution is needed because Hungarian constituent order is strongly influenced by information structure, while English has relatively fixed ordering. Therefore, the relevant monolingual structures should be established before treating a boundary to be equivalent. A structural compatibility of potential switch sites can be observed in Example 15 (see Table 7), comparing Hungarian monolingual structure to English:

Hungarian bilingual phrase: *'láttátok a progress-t'*

Hungarian: *láttátok a fejlődést* English: *you saw the progress*

Despite its considerable influence, the universality of the Equivalence Constraint has been questioned, most explicitly by Berk-Seligson (1986), whose Spanish-Hebrew data included switches incompatible with the proposed surface-equivalence requirement. The constraint is therefore retained here as a foundational component of the Poplackian approach.

4.3.2 The Free Morpheme Constraint

The Free Morpheme Constraint restricts code-switching between a bound morpheme and a lexical stem from another language (Poplack, 1980; Sankoff & Poplack, 1981), so it can be directly linked to intraword language mixing. Importantly, the original account distinguishes such switching from forms in which the lexical material has been integrated into the language of the bound morphology.

This qualification is central to Hungarian-English forms consisting of an English-origin base and Hungarian affixation. Their mixed etymological composition does not establish that switching has occurred across a morpheme boundary. If the English-origin item has been

incorporated into Hungarian as a borrowing, the attachment of Hungarian morphology is not analysed as a switch between English and Hungarian and it does not violate the Free Morpheme Constraint. Based on this, the main claim of this constraint is that language origin and contact process are not equivalent. An English-origin stem followed by Hungarian morphology should be evaluated for its grammatical behaviour before it can be treated either as an intraword code-switch or a recipient-language-integrated borrowing.

4.3.3 Code-Switching and Borrowing

Distinguishing between code-switching and borrowing became a central concern of Poplack's work because lexical origin alone cannot identify the process responsible for an other-language item in bilingual discourse. This problem is especially relevant to lone items: unlike a multiword constituent, an isolated lexical element contains little to no internal syntax so as to infer the active grammatical system.

For this reason, Poplack and colleagues analyse donor-language-origin material through its linguistic behaviour and social distribution (Poplack & Sankoff, 1984; Poplack et al., 1988, 1989). Code-switches are identified as longer, multiword elements that showcase donor-language grammatical organization, whereas lone L^D items characteristically pattern with recipient-language morphosyntax. The earlier mentioned Finnish-English study of Poplack et al. (1989) is especially relevant as it demonstrates this distinction in contact by referring to *kidney-stä* 'kidney.N-ELAT' as a borrowing while in the clause *yliopisto-ssa otti, niin kuin, art history* (tr.: at university she took, like, art history') *'art history'* as a code-switch.

However, the distinction between lone and multiword material should not, be reduced to constituent length according to Treffers-Daller (2025), who proposes the Simple View of borrowing and code-switching. Here, the central criterion for borrowing is listedness, which means that material stored as a lexical unit in the recipient-language speaker's lexicon constitutes borrowing, whereas material generated online, through productive donor-language grammar, constitutes code-switching. On this account, multiword units can also be borrowings if they are stored and retrieved as units rather than productively assembled during bilingual production. While this view could be applied to Hungarian-English data, the lexicon of the speaker is not accessible because of the nature of the corpus, therefore testing it would be problematic for this study.

Nevertheless, morphosyntactic integration must be distinguished from social establishment. According to Poplack (2018), an item may be grammatically accommodated to the recipient language without being recurrent or community-wide, while an established loan is socially distributed in addition to being part of recipient-language grammar. She also notes that phonological adaptation, apparent foreignness, frequency, or dictionary attestation cannot individually determine the distinction; grammatical patterning must be considered together with distributional evidence.

4.3.4 Established and Nonce Borrowing

The concept of nonce borrowing accounts for donor-language-origin items that behave grammatically like recipient-language vocabulary despite lacking the recurrence and community distribution of established loans. Sankoff et al. (1990) distinguish such forms from established borrowings principally through their social and distributional status, not through an assumed difference in morphosyntactic integration.

Therefore, they reject a necessary developmental sequence from an unintegrated code-switch to becoming an integrated borrowing. A lexical item may instead be borrowed spontaneously and immediately participate in recipient-language grammatical processes. Thus, an item occurring only once cannot be classified as a code-switch only because it is not frequent, just as dictionary absence does not establish code-switch status.

The distinction is relevant to the present Hungarian-English corpus because English-origin items vary considerably in observable conventionalisation. However, corpus-internal recurrence alone cannot establish whether an item is genuinely nonce or only community-wide, so any claim about established versus nonce status must remain proportional to the available evidence concerning frequency, distribution across speakers, dictionary attestation, and broader Hungarian usage.

4.3.5 The Nonce Borrowing Hypothesis

The Nonce Borrowing Hypothesis (NBH) formalises the prediction that spontaneously borrowed lexical items can display recipient-language morphosyntax right from their first use, before becoming socially established (Poplack et al., 1988; Sankoff et al., 1990). Nonce and established borrowings are therefore expected to pattern grammatically like recipient-language lexical items despite differing in frequency and community distribution. Diachronic evidence

from Poplack and Dion (2012) supports this position: donor-language items displayed recipient-language linguistic integration from their earliest observations and not gradually acquired it as their frequency increased.

Poplack (2012) also clarifies that the NBH is not a developmental hypothesis according to which nonce borrowings necessarily become established loans, nor does it predict that increasing frequency comes along with progressive grammatical integration. According to her, an item may immediately participate in recipient-language morphology and syntax even if it remains rare or restricted to an individual speaker.

Since recipient-language patterning does not require overt morphology in every case, the appropriate test is to compare L^D-origin items with L^R items that should exhibit the same grammatical accommodation. The emphasis of Poplack and Meechan (1998) is on the comparison with native forms in unmixed L^R discourse. Accordingly, the presence of a Hungarian suffix on an English-origin item may demonstrate morphosyntactic integration, but it does not constitute a full test of the NBH.

The strongest evidence for the NBH should probably come from low-frequency English-origin items that pattern according to the same Hungarian grammatical requirements as established loanwords and comparable Hungarian lexical items. In conclusion, the occurrence of Hungarian morphology on lone English bases can be treated as compatible with nonce borrowing, but still not as conclusive evidence for the hypothesis.

4.3.6 The Comparative Method and Recipient-Language Benchmarking

The methodological basis of Poplack's (2012, 2018) later work is the comparative variationist method, which evaluates contact forms against independently established grammatical patterns instead of classifying them by isolated surface characteristics. Where the data permit, the comparison involves native recipient-language items, lone L^D-origin items, established borrowings, and independently identifiable code-switches.

This means that the analysis of this approach is distributional rather than simply morphological. If a grammatical feature is variable in native L^R speech, integration requires L^D-origin items to behave the same way. In this way, borrowed material is prevented from being

held to a stronger grammatical standard than native vocabulary and the empirical assessment of L^R-equivalent behaviour is allowed.

At the same time, such a benchmarking requirement limits what can be concluded from the present corpus as it was not designed as a community-wide variationist sample which should contain matched sets of native Hungarian forms, established loans, nonce borrowings, and code-switches in every grammatical environment.

4.3.7 Criticisms and Methodological Limitations

The strength of the approach in discussion also poses a rather big methodological demand. Identifying a form as indeed nonce requires evidence of its recurrence and community distribution, while the establishment of L^R-equivalent behaviour requires sufficient comparable native, borrowed, and switched material.

This is quite important for the present podcast corpus: recurrence across speakers, external corpus frequency, and dictionary attestation can provide useful evidence concerning the establishment of the items, but they are not identical to the community-distribution evidence which is a feature of the original variationist methodology. Also, an item occurring once in the corpus cannot automatically be assumed to be nonce, and widespread dictionary or corpus attestation need to be treated as evidence of establishment and not of its sociolinguistic status in the sampled community.

Concerning the status of LOLIs, Backus (2014) questions the assumption that borrowing and code-switching must always be mutually exclusive. In his usage-based account, code-switching is synchronic, whereas borrowing is a diachronic process of entrenchment and conventionalisation. According to him, the same other-language item may therefore function as a switch in a certain bilingual utterance while already being partly entrenched in an individual's usage or conventionalised within the speech community. Also, Treffers-Daller (2025) challenges the integration-based classification and argues that listedness provides a stronger classification of borrowing than morphosyntactic integration.

Briefly mentioned in Chapter 3, a direct empirical challenge to the NBH is provided by Stammers and Deuchar (2012), where they examined the application of Welsh soft mutation to English-origin verbs, treating mutation in grammatically appropriate environments as a measure of integration into Welsh. Their results showed that more frequent English-origin verbs underwent mutation more often, while less frequent and dictionary-unlisted English-origin

verbs showed lower rates of mutation. They interpret this pattern as inconsistent with the prediction that nonce borrowings should display L^R integration comparable to established loans. As a conclusion, their findings suggest that, at least, some aspects of grammatical integration may be gradient and frequency-sensitive rather than immediate.

4.4 Comparison of the Matrix Language Frame and Poplackian Variationist Approaches

Having outlined the two frameworks separately, now the focus is on finding the points where they make overlapping or different predictions about the same Hungarian-English constructions. Although both approaches treat the grammatical behaviour of other-language material as systematic, they differ in how it is represented and what conclusions can be drawn from morphosyntactic integration.

4.4.1 Shared Assumptions and Areas of Convergence

The two approaches share three assumptions that make their comparison possible. First, neither treats bilingual constructions as grammatically arbitrary. The MLF Model captures their systematicity through structural asymmetry and the constraints associated with the Matrix Language (Myers-Scotton, 1993, 2002), while the Poplackian approach establishes grammatical patterning through comparison with recipient-language material and independently identifiable code-switches (Poplack, 2018).

Second, both distinguish lone lexical material from larger constituents with identifiable internal structure. In the MLF Model, an individual EL content morpheme in an ML+EL constituent differs from an EL island organized internally according to Embedded Language grammar (Myers-Scotton, 1993, 2002). Poplack likewise distinguishes lone L^D-origin items from multiword code-switches that retain donor-language grammatical organization (Poplack, 2018).

Third, according to both views, lexical origin alone is an insufficient diagnostic for analysis. An English-origin form must be examined in its grammatical environment before its framework-specific classification. This is consistent with the theory-neutral distinction made in Chapter 3 among LOLIs, intraword mixed forms, mixed constituents, mixed compounds, and internally English multiword constituents.

4.4.2 Matrix Language, Recipient Language, and the Interpretation of Integration

Undoubtedly, the terms Matrix- and recipient language refer to different theoretical constructs. Again, in the MLF Model, the ML provides the morphosyntactic frame of a bilingual CP, while into which EL material is integrated (Myers-Scotton, 1993, 2002) and in the Poplackian approach, the L^R is the language into which L^D material is incorporated and question is whether the L^D-origin item follows independently established L^R grammatical patterns (Poplack, 2018).

As this parallel is apparent in the constructions examined in this study, both views may identify Hungarian as the grammar-providing language, so would label Hungarian as ML and L^R accordingly.

At the same time, this overlap explains why the presence of integration on its own cannot discriminate between the frameworks. For example, English-origin nouns carrying Hungarian case morphology, may be consistent with English EL content morphemes occurring in a Hungarian ML frame, while it can also provide evidence of L^R integration of L^D items within the Poplackian analysis.

Thus, the difference lies in the predictions of additional claims of the two views. According to the MLF, the construction must conform to the structural predictions associated with ML control, including the MOP and the 4-M-informed SMP, and for the Poplackian approach, a stronger borrowing analysis requires the English-origin form to pattern like comparable Hungarian lexical material under equal grammatical conditions.

4.4.3 The Status of Lone and Internally Structured Other-Language Material

LOLIs provide a point of disagreement in classification. Under the MLF Model, an English noun, verb, or adjective may occur as an EL content morpheme in a Hungarian ML+EL constituent (Myers-Scotton, 1993, 2002). A form such as *content-et* can be analysed as English lexical material occurring within a Hungarian ML frame; however, the same form supports Poplack's (2018) borrowing analysis if *content* follows the same case-marking and syntactic patterns as comparable Hungarian nouns, here *tartalm-at* 'content.N-ACC'.

Not surprisingly, the presence of Hungarian morphology on lone items is compatible with both frameworks; the difference is whether the English-origin stem remains EL material or is analysed as part of Hungarian lexical usage.

Multiword English-origin material is an even more sensitive point of interest. A sequence containing several English-origin words is not necessarily a syntactic code-switch. It may also be a compound, a fixed expression, or a complex lexical unit. Only material with independently identifiable English internal grammatical organization provides strong evidence for classifying it as an EL-island under the MLF or L^D code-switching in the Poplackian approach.

4.4.4 Frequency, Establishment, and Community Evidence

Frequency carries different evidential value in the two frameworks. In the MLF Model, the frequency or conventionalization of an English-origin lexical item should not determine whether a bilingual construction satisfies the MOP or the 4-M-revised SMP. These predictions concern the structural organization of the bilingual constituent, so a rare and an established English-origin item can provide equal evidence for or against an MLF prediction if they occur in the same grammatical configuration.

This, however, is more important for the Poplackian view as evidence concerning lexical establishment, but not as a degree of morphosyntactic integration. The NBH predicts that a spontaneously borrowed item may display L^R grammar before becoming frequent or socially widespread (Sankoff et al., 1990; Poplack, 2012). Poplack and Dion's (2012) findings must be mentioned again here, as the integration of L^D was already present in early attestations, whereas frequency only distinguished items that became established from those that did not.

Mixed constructions can be tested against an MLF structural prediction without establishing how widely its English-origin lexical item is used in the speech community; however, claiming that the same item is an established or nonce borrowing would require evidence beyond its internal morphosyntax (e.g. recurrence across speakers, broader corpus frequency, and dictionary attestation), yet individually, they still do not establish a concrete community status (Poplack, 2018).

Applying these ideas for the present corpus, frequency cannot be used to infer the degree of Hungarian grammatical integration, and a single occurrence of an English-origin item is not automatically classified as nonce. Instead, the morphosyntactic behaviour is compared to determine their according classification or the lack thereof.

4.4.5 Evidential Requirements, Falsifiability, and Model Comparison

Comparing the frameworks requires a distinction between compatibility and diagnostic evidence. A construction is compatible with a framework when it does not violate its predictions; it is diagnostically informative only when the framework makes a prediction that excludes a plausible alternative. This distinction is especially important where Hungarian and English independently permit the same structure.

Therefore, the strongest tests for the MLF are grammatical environments in which Hungarian and English differ. Hungarian word order supports the MOP most clearly where the equivalent English structure would require a different order. Similarly, Hungarian morphology provides a strong test of the SMP only when the relevant morpheme can be identified as an outsider late system morpheme. A Hungarian suffix whose 4-M status is uncertain cannot be counted as equally strong evidence, although still can qualify as compatible with the model. Nguyen (2025) demonstrates the importance of this distinction in the extensively overlapping word order of Vietnamese and English, as well as in examples where limited overt morphology left many constructions without a clear identification a ML reliably.

Potential counter-evidence to MLF must also be defined narrowly. Once Hungarian has been established independently as the Matrix Language, English ordering in an ML+EL constituent would challenge the MOP, and likewise, an English outsider late system morpheme in a ML+EL constituent would challenge the 4-M-informed SMP. Yet, neither case is treated as counter-evidence if the construction is independently shown to be an EL island, being classified by their structure. These requirements address the concern about the falsifiability of the MLF Model raised by MacSwan (2005) and defended against by Jake et al. (2005).

For the NBH, besides testing whether the L^D item follows the same grammatical conditioning as comparable Hungarian forms, a further condition must be satisfied: the item used to test the hypothesis must have independent evidence of being socially unestablished. Otherwise, recipient-language integration can be demonstrated, but the specific prediction classifying nonce borrowings cannot be tested. Counter-evidence to the NBH would require nonce borrowings to pattern systematically differently from appropriate Hungarian or established-loans. As Stammers and Deuchar (2012) show in their Welsh soft-mutation findings (where rates of mutation differed according to lexical frequency), such a challenge has been presented already.

Thus, the two frameworks can be compared both on their predictive accuracy and analytical coverage. A construction may be consistent with a framework, inconsistent with one of its predictions, or indeterminate because the framework does not provide sufficient criteria for resolving its status. These outcomes are not equivalent: an indeterminate case does not constitute direct counter-evidence, but a framework that produces substantially more indeterminate analyses across the same dataset has lower explanatory coverage.

On this basis and following Duran Eppler et al. (2017) and Deuchar and Stammers (2016), the analysis of the data distinguishes three possible outcomes for each prediction: consistent, inconsistent, and indeterminate. Consistent cases contribute to the framework's empirical support, inconsistent cases identify potential counter-evidence, and indeterminate cases indicate limitations in analytical coverage under the available criteria. In conclusion, the proportion and distribution of all three outcomes can be compared across the two frameworks.

4.5 Predictions for Hungarian-English Morphosyntactic Integration

The preceding sections establish the theoretical criteria against which the Hungarian-English data can be evaluated. Now, the predictions below are formulated only for properties that can be identified in the corpus and operationalised later. They concern Matrix Language identification, morpheme order, system-morpheme source, the treatment of internally English material, recipient-language grammatical patterning, lexical establishment, and the status of multiword constructions.

4.5.1 Predictions of the Matrix Language Frame Model

The MLF framework makes five main predictions for the Hungarian-English data. First, Hungarian should function as the Matrix Language in the intraclausal constructions. Its status should be identifiable from clause-level structural evidence, including predicates, function words, argument structure, constituent organization, and grammatical morphology, and not from the number of Hungarian words (Myers-Scotton, 1993, 2002).

Second, under the Morpheme Order Principle, ML+EL constituents should follow Hungarian order where Hungarian and English make different predictions (Myers-Scotton, 1993, 2002), which includes the ordering of morphemes in mixed words and specifically Hungarian syntactic patterns such as coverb placement. Structures compatible with both languages are predicted to be possible but are not clear diagnostic tests of the MOP.

Third, as for the 4-M-informed System Morpheme Principle, outsider late system morphemes in Hungarian ML+EL constituents should come from Hungarian. English outsider late system morphology in such a constituent would be inconsistent with this prediction unless it has formed as part of an identifiable EL island. Early and bridge system morphemes are not necessarily subject to this restriction and must be distinguished from outsider late system morphemes.

The fourth prediction states that English-origin lexical material should conform to the grammatical requirements of the Hungarian frame. So, English-origin verbal bases are expected to undergo the category formation and inflection required by Hungarian, while nominal and adjectival material should receive Hungarian grammatical morphology where and if the relevant construction requires it. English-origin verbal material is consequently expected to occur with Hungarian rather than English finite grammatical morphology in ML+EL predicates (Bolonyai, 2005; Myers-Scotton & Jake, 2014).

Lastly, multiword material with independently identifiable English internal grammar may be treated as an EL island and it can retain English morpheme order and grammatical morphology (Myers-Scotton, 1993, 2002), while its external grammatical realization is supplied by Hungarian. The Model predicts Hungarian control of the external morphosyntactic frame, Hungarian ordering at diagnostic conflict sites, Hungarian outsider late system morphology, and English internal grammar in EL islands.

4.5.2 Predictions of the Poplackian Approach and the Nonce Borrowing Hypothesis

The Poplackian analysis makes three principal predictions for the same data. First, lone English-origin items functioning as borrowings should follow Hungarian morphosyntactic patterns under the same conditions as comparable Hungarian items (Poplack & Sankoff, 1984; Poplack et al., 1988, 1989; Poplack & Meechan, 1998; Poplack, 2018). Relevant evidence includes case, number, possession, verbalization, derivation, and syntactic distribution. While overt Hungarian morphology is strong evidence of integration where Hungarian requires such marking, its absence is not evidence of the absence of integration where comparable Hungarian forms would also be unmarked.

Second, the Nonce Borrowing Hypothesis predicts that social establishment is not required for Hungarian grammatical integration (Sankoff et al., 1990; Poplack, 2012; Poplack & Dion,

2012). An English-origin item for which non-establishment can be supported should therefore be capable of displaying the same Hungarian morphosyntactic behaviour as established borrowing. Frequency or dictionary non-attestation alone is insufficient to establish nonce status. Where reliable evidence of establishment is unavailable, the item cannot be used as a definitive test of the NBH.

Third, productively structured multiword code-switches should retain identifiable English internal grammar (Poplack, 2018). English function words, English inflection, coordination, or other English-specific internal dependencies provide evidence for this classification. The occurrence of several English-origin words alone is insufficient, because compounds and lexicalized expressions may not represent productive multiword switching. Multiword constructions whose internal status cannot be established, particularly those that also receive Hungarian external morphology remain indeterminate.

Thus, the Poplackian approach predicts Hungarian grammatical behaviour for borrowing-compatible lone items, immediate recipient-language integration where nonce status can be established, and retention of English internal grammar in identifiable multiword code-switches.

4.5.3 Summary of Framework-Specific Predictions

Table 10 summarizes the main predictions of the two approaches for Hungarian-English configurations examined in the dissertation. These will be operationalized in later chapters.

Table 10. Summary of the MLF and the Poplackian frameworks' predictions

Domain	MLF prediction	Poplackian / NBH prediction
Clause-level grammatical organization	Hungarian should be identifiable as the Matrix Language from clause-level structural evidence.	Hungarian provides the L^R-grammar against which English-origin material is compared.
Lone English-origin items in Hungarian structure	They may occur within a Hungarian ML+EL constituent if the construction satisfies Hungarian ML requirements.	They are treated as borrowings and should pattern morphosyntactically like Hungarian items under the same conditions.

Morpheme and constituent order	ML+EL constituents should follow Hungarian order where Hungarian and English make different predictions, but structures compatible with both languages are allowed.	L^R ordering contributes to a borrowing analysis where a lone English-origin item follows the same structural pattern as comparable Hungarian, while L^D order is treated as CS.
Outsider late system morphology	They should come from Hungarian. English outsider late system morphology would be inconsistent with the prediction unless it occurs within an EL island.	Hungarian grammatical morphology is predicted for borrowings on the same conditions as comparable Hungarian forms; otherwise treated as CS.
Category formation and inflection	English-origin verbal bases should receive the Hungarian category formation and inflection required by the predicate structure; comparable accommodation is expected in nominal and adjectival domains.	Hungarian case, number, possession, verbalization, derivation, inflection, or zero marking support borrowing-compatibility, reproducing the grammatical behaviour of comparable Hungarian items.
Socially unestablished / nonce items	Frequency and social establishment do not determine whether a construction satisfies the predictions of the Model.	Nonce or socially unestablished borrowings may display Hungarian morphosyntax.
Internally English multiword material	Such forms are treated as EL islands and retain English morpheme order and grammatical morphology.	Code-switches should retain identifiable English internal grammar, as multiword status alone is insufficient.
Multiword English material + Hun. morphology	Complex English-origin material may remain part of an ML+EL construction.	If the English sequence cannot be established as a CS or a lexical unit, its status is indeterminate.

5. DATA AND METHODS

5.1 Research Design

The present study adopts a corpus-based, analytical design to investigate morphosyntactic integration in naturalistic Hungarian-English bilingual speech. The data consist of publicly available Hungarian-language podcast recordings produced independently of the research. After their initial selection, the bilingual utterances, treated as units of analysis, were manually extracted from the podcasts recordings and were transcribed and analysed in both frameworks.

First, English-origin material is coded according to observable structural properties without assigning an immediate code-switching or borrowing classification. These properties include constituent size, word class, Hungarian derivational and inflectional morphology, case and possessive marking, verbalization, coverb use, constituent structure, and the internal organization of multiword English material. This theory-neutral coding provides a common empirical basis for the later framework-specific analyses. Later, the coded constructions are evaluated separately under the MLF Model and the Poplackian approach and their predictions introduced in the previous chapter.

Framework-level coverage and prediction-specific evidence are handled separately. Individual predictions are tested only in constructions that provide the grammatical or distributional evidence required for that test; non-diagnostic structures are not counted as equal confirmations. At the broader framework level, constructions may receive a determinate analysis, conflict with an applicable prediction, or remain indeterminate where the framework or available information does not permit a reliable analysis. This design makes it possible to compare both the empirical support for individual predictions and the analytical coverage of the two frameworks while keeping overt morphosyntactic integration separate from a specific interpretation.

5.2 Data Source and Corpus

The dataset consists of naturally occurring Hungarian-English bilingual speech collected from publicly available Hungarian-language podcasts uploaded to YouTube. It contains approximately 30 hours of video-recorded material from 24 podcast episodes and includes 37 speakers between 20 and 69 years of age (at the time of recording). In all cases, Hungarian is the first and dominant language of the speakers. Based on their observable use of English in the

recordings, their proficiency was estimated to range broadly from A2 to C1 level; no standardized proficiency assessment was available due to the nature of the data. These recordings represent mostly informal and semi-formal conversational settings.

Podcasts were selected because they provide extended stretches of relatively spontaneous interaction, while they preserve sufficient audio- and video quality for detailed linguistic analysis. Videos are especially important to have access to, since they can help in distinguishing between possible ambiguities or less audible segments by providing an option for lip reading. The material was not produced for the purposes of the present research, and the speakers were not responding to linguistic elicitation tasks, which strengthens its natural characteristic. At the same time, it cannot be treated as entirely unmediated spontaneous conversation, as individual episodes may differ in preparation, editing, topic structure, audience orientation, and the degree of formality. Therefore, the corpus is treated as a source of naturalistically occurring, non-elicited bilingual speech, while not claiming to be equivalent of private, everyday conversation.

The initial corpus contained 407 utterances involving English-origin material. For the morphosyntactic analyses, tokens were further filtered based on criteria specified in the following sections, including constituent type, availability of relevant morphological or syntactic evidence, and eligibility for framework-specific testing. Utterances containing monolingual English clauses were excluded from the final dataset, as the current study focuses on intraclausal language mixing. The data are shared online and can be freely accessed through the links in the Appendices, providing recheckability and reproducibility features for this study.

All recordings were accessed from a publicly available source, by uploading onto which, speakers agree to share that content with anyone, including researchers. Although the speakers are well-known personalities in the Hungarian community, their identities are not relevant to the linguistic analysis, and as Stommel and de Rijk (2021) demonstrate, research based on publicly available online material still requires consideration of identifiability, retrievability, and potential harm; thus, speaker identities are anonymised during the presentation of individual examples.

5.3 Podcast Selection Criteria

Episode selection was not restricted by topic, profession, or communicative domain, and no particular English-origin lexical items or grammatical constructions were targeted in advance.

An episode was included if it contained predominantly Hungarian conversational speech produced by native and dominant speakers of Hungarian, together with any identifiable English-origin material. Unscripted or minimally scripted interactions were of preference. Episodes with clearly scripted performances, and recordings of poor audio or video quality which did not permit reliable transcription were excluded. Proper recording quality was required for identifying lexical material, morphological boundaries, and potentially ambiguous segments. Within these criteria, episodes were selected without favouring certain structural types of bilingual construction.

Only publicly accessible recordings were included, and no private or restricted-access material or additional personal information about the speakers was collected. However, public availability was not treated as equivalent to explicit research consent; and that is the main reason behind anonymisation. This procedure follows the principle that public accessibility does not remove the researcher's responsibility to consider contextual integrity and participant identifiability arising from the reproduction or analysis of online material (British Association for Applied Linguistics [BAAL], 2021).

5.4 Identification of Bilingual Constructions and Units of Analysis

The initial identification procedure was kept separate from the later framework-specific classification of the data because a lone English-origin item may be analysed as an Embedded Language element in the MLF Model or as borrowing within the Poplackian approach. The purpose of this stage was to establish a reproducible set of potential Hungarian-English candidates, not to determine their final theoretical status.

A dictionary criterion was used to exclude clearly conventionalised Hungarian loans from the dataset. English-origin forms listed as established Hungarian lexical items in *Magyar értelmező kéziszótár* [*Concise Explanatory Dictionary of Hungarian*] (Akadémiai Kiadó, n.d.) were excluded. Forms not listed in the reference source were provisionally retained where their English origin was otherwise identifiable. This procedure follows the operational use of dictionary evidence in previous corpus-based code-switching research (Deuchar, 2006).

As a form may be absent from a general dictionary while being conventionalised within particular social, professional, or online communities, dictionary non-attestation was not interpreted as proof of code-switching or nonce status. On the other hand, a single occurrence in the corpus cannot establish that an item is novel. Thus, dictionary status therefore served

only as an initial filtering criterion, and possible borrowing status was reconsidered during the later Poplackian analysis (Poplack, 2012, 2018; Sankoff et al., 1990).

Proper nouns, geographical names, brand names, and titles functioning as names were not counted towards mixed constructions. Discourse markers were retained where they were clearly identifiable as English-origin forms and were not established Hungarian loans according to the same criterion. Items formally compatible with both languages (e.g. *video/videó*) were evaluated using their meaning, pronunciation, grammatical environment, and discourse context. If language affiliation was unresolved (in the case of *platform*, for example), the token was not treated as English-origin for the analyses.

Interclausal English material was also excluded from the morphosyntactic dataset because an internally English complete clause does not provide sufficient evidence of grammatical interaction between Hungarian and English. Therefore, the resulting corpus only contained intraclausal constructions in which the two languages could potentially interact.

5.5 Transcription and Glossing

All utterances containing an identified switching candidate were transcribed manually. The surrounding utterance was retained because clause-level context was necessary for determining grammatical function, constituent structure, ML evidence, and L^R requirements. The individual English-origin item or bilingual construction within that context was then subjected to detailed structural coding.

English-origin lexical material was distinguished from Hungarian grammatical material, and morphologically complex forms were segmented into their identifiable components. Pronunciation was considered where necessary for resolving language affiliation or morphophonological form, but the corpus was not subjected to complete phonetic transcription.

Examples requiring morphological analysis were presented using interlinear morpheme-by-morpheme glosses based on the Leipzig Glossing Rules (Comrie et al., 2008). Hyphens mark boundaries between segmentable morphemes, and grammatical functions are represented by standardized abbreviations (see list of glosses in the appendices). Each cited example contains the original bilingual form, a grammatical gloss where relevant, and an idiomatic English translation.

5.6 Coding and Analytical Classification

The primary coding unit was the individual English-origin item or bilingual construction, however, the transcribed clause in which it occurred was also glossed and retained as contextual evidence for determining the grammatical function and structure of the unit. Morphologically complex forms were segmented into their English and Hungarian components.

Each construction was coded descriptively. The variables recorded were the following:

- lexical category of the English-origin material;
- constituent type and size;
- presence and type of Hungarian derivational or inflectional morphology;
- case, number, possessive, verbal, adjectival, and degree marking where applicable;
- syntactic environment and grammatical function;
- presence and position of Hungarian coverbs;
- internal organization of multiword material; and
- language source of identifiable lexical and grammatical morphemes.

This initial stage, again, distinguished observable structural properties from theoretical interpretations such as EL insertion, code-switching, established borrowing, or nonce borrowing.

For the MLF analysis, additional coding recorded the identified Matrix Language, whether the construction provided a diagnostic test of morpheme order, the language source and 4-M classification of relevant system morphemes, and whether multiword English material satisfied the criteria for an EL island. Structures in which Hungarian and English permitted the same ordering were treated as non-diagnostic for the MOP.

For the Poplackian analysis, lone English-origin items were coded according to whether their grammatical behaviour corresponded to that expected of Hungarian material in the same environment. Evidence concerning lexical establishment across speakers was not available, therefore, morphosyntactic integration was operationalised as a measure of borrowing or CS status, where the structure of the construction allowed. Multiword material was coded according to whether identifiable English internal grammatical organization could be established.

Morphologically integrated multiword constructions whose status could not be resolved within the canonical lone-item versus multiword-switch distinction were retained as indeterminate.

As mentioned in Chapter 4, for each applicable theoretical prediction, the analytical outcome was coded as consistent/accounted for, inconsistent/not accounted for, or indeterminate:

Consistent indicating that the available evidence satisfied the relevant prediction;

Inconsistent where there is a conflict with a prediction without an independently justified exception; and

Indeterminate indicating that the structural or distributional evidence required for classification was insufficient.

5.7 Quantitative Analysis

This quantitative analysis of the data was based on frequency counts and proportions calculated for the structural categories established during coding. Since several categories overlap, the denominator used for each calculation was determined separately.

The initial dataset included a total of 419 English-origin occurrences within 407 bilingual utterances. Of these, 405 were intraclausal and 14 interclausal. Only the intraclausal occurrences were retained for the later morphosyntactic analyses.

The 405 intraclausal constructions consisted of 302 lone or single-item insertions and 103 multiword constructions, representing 74.6% and 25.4% of the intraclausal dataset, respectively. Within the intraclausal set, a total of 187 constructions contained an intraword language boundary as well.

It is important to note that intraword status was treated as an overlapping structural property rather than as a separate constituent-size category, since of these constructions, 152 involved LOLIs and 35 involved multiword English material carrying Hungarian morphology. The remaining dataset consisted of 150 lone items and 68 multiword constructions without intraword morphological integration.

6. RESULTS

6.1 Overview of the Corpus

As stated above, the corpus contained 407 bilingual utterances, within which 419 individual English-origin occurrences were identified. After excluding the 14 occurrences with interclausal switches the resulting analytical dataset consisted of 405 intraclausal constructions occurring within 393 bilingual utterances.

Of these, 302 (74.6%) were lone or single-item constructions and 103 (25.4%) were multiword constructions. English-origin material occurred as nouns, adjectives, verbal bases, adverbs, discourse markers, compounds, fixed expressions, and larger nominal or other sequences.

Intraword integration occurred in 187 constructions (46.2%). Of these, 152 were lone items and 35 were multiword constructions carrying Hungarian morphology. The remaining dataset consisted of 150 lone items without intraword morphology and 68 multiword constructions without intraword morphology. The four resulting structural categories, based on intraword integration and the number of English words of the insertions, were the following:

- lone, non-intraword constructions: 150 (37.0%);
- lone intraword constructions: 152 (37.5%);
- multiword, non-intraword constructions: 68 (16.8%);
- multiword intraword constructions: 35 (8.6%).

Thus, constituent size and intraword integration were treated as independent dimensions. While most constructions contained only one English-origin lexical item, Hungarian morphology also occurred on 35 of the 103 multiword constructions (34.0%).

6.2 Morphosyntactic Integration Patterns

Hungarian morphosyntactic integration was found across nominal, verbal, and adjectival domains. The attested processes included case and number marking, possession, verbalization, verbal inflection, adjectival derivation, degree marking, nominalization, and the incorporation of English-origin verbal material into Hungarian coverb constructions.

6.2.1 Nominal Integration

English-origin nouns occurred both with and without overt Hungarian morphology. Morphologically unmarked forms included examples like *content*, *expert*, *update*, *shop*, and *output*. These items occupied nominal positions in Hungarian clauses and could combine with Hungarian determiners, predicates, and other surrounding grammatical material without receiving an overt suffix as the Hungarian grammar did not require one.

Overtly marked nouns received Hungarian nominal morphology according to their grammatical environment. Examples include *content-et*, 'content.N-ACC' *badge-et*, 'badge.N-ACC' *game-ről* 'game.N-DELAT', *mainstream-ből* 'mainstream.N-ELAT', *fullscreen-be* 'fullscreen.N-ILLAT', *reset-ként* 'reset.n-ESS', *shop-ba* 'shop.N-ILLAT', and *rush-nál* 'rush.N-ADESS'.

Nominal morphology was also attached to multiword English material. Such examples are presented in Table 11.

Table 11. Multiword English material with Hungarian nominal morphology

Ex.22	*'community guideline set-tel'*
Gl.:	community.N guideline.N set.N-INSTR
Ex.23	*sleep tracking-re*
Gl.:	sleep.N tracking.N-SUBLAT

In such forms, the Hungarian suffix occurred on the final element of the English sequence and expressed the grammatical relation of the complete nominal construction.

A structurally different group consisted of mixed English-Hungarian nominal constructions in which an English-origin modifier occurred with a Hungarian head, as in *review videó (video)*, *win-win helyzet (situation)*, *content gyártás (production)*, *mental health téma (topic)*, *viewing élmény (experience)*, *donate gomb (button)*, and *believe felirat (label)*. When these constructions received Hungarian inflection, morphology occurred on the Hungarian head, as in *review videó-k-at* (video.N-PL-ACC) and *win-win helyzet-nek* (situation.N-DAT).

6.2.2 Verbal Integration

The corpus contained 40 constructions involving English-origin material integrated by Hungarian verbal morphology, giving 9.9% of the intraclausal dataset and 21.4% of the intraword constructions.

All 40 contained a Hungarian verbalizer suffix either the *-l* or *-z* with their phonologically conditioned linking vowels. Such examples are observable in Table 12.

Table 12. Examples of Hungarian verbalizer use on English items

	Mixed items	Glossing
Ex.24	*force-**ol**-ják*	force.N-VBZ-3PL.DEF
Ex.25	*spoiler-**ez**-z-em*	spoiler.N-VBZ-IMP-1S.DEF
Ex.26	*roadshow-z-ik*	roadshow.N-VBZ-3S.INDEF

The verbalized bases subsequently received the morphology required by the Hungarian morphosyntactic environment. Person and number occurred in forms such as *stream-el-**ünk*** and *stream-el-**sz***; definite conjugation in *force-ol-**ják*** and *hijack-el-t-**ék***; conditional morphology in *discount-ol-**nám***; potential morphology in *flex-el-**het**-ek*; and imperative morphology in *jam-el-**j**-ünk*. Non-finite forms included examples like *cancel-öl-**ni***, *spoiler-ez-**ni***, and *fingerflip-el-**ni***.

Further derivation could follow verbalization in forms such as *gameplay-ez-**és**-nél*, *flex-el-**és**-ről*, and *flex-el-**és**-nek*. These contained a Hungarian verbalizer followed by a nominalizer morpheme and, where needed, case marking afterwards.

Verbalization also applied to multiword English bases, including *trash talk-**ol**-t*, *new age-**el**-t-ek*, and *face app-**oz**-ott*. This means that the base selected for Hungarian verbal derivation was not restricted to a single English word.

6.2.3 Adjectival Integration

English-origin adjectival material occurred both without overt morphology and with Hungarian derivational or degree morphology. Such bare forms included *cringe*, *clean*, *deep*, *viral*, *toxic*, *smooth*, *relatable*, and *up-to-date* in Hungarian attributive or predicative environments.

Hungarian *-s* and its variants (with linking vowels) occurred with English-origin bases in forms like *full-os*, *top-p**os**-ak*, *pick-up truck-**os***, *sales-**es***, *stand-up-**os***, and *late-night show-**s***. Depending on the construction, these forms functioned either adjectivally or as nouns referring to things or persons associated with the relevant activity or domain.

Hungarian comparative and superlative morphemes were also productive (see Table 6), especially in Example 13, ***leg**-light-os-**abb**-ak.* This construction contained several layers of Hungarian adjectival morphology: the superlative prefix *leg-*, adjectival derivation, comparative marking, and it also received the plural suffix.

Other English-origin adjectival material received morphology determined by its syntactic environment, as in *cool-nak*, *real-nek*, and *old-school-ak*. Further category formation occurred in *happy-ség*, where the English adjective served as the base for a Hungarian abstract noun. Table 13 presents examples of these in their syntactic environment.

Table 13. Adjectival integration in Hungarian morphosyntax

Ex.27	*'aki-k-et … nagyon* real*-nek tart-ok'* (tr.: *those whom … I find very real*)
Gl.:	who.PRON.INTER-PL-ACC very.ADV real.ADJ-DAT keep.V-1S
Ex.28	*'ami-k tényleg ilyen* old-school*-ak'* (tr.: *which are old-shool like this*)
Gl.:	what.PRON.INTER-PL really.ADV such.ADJ old-school.ADJ-PL

6.2.4 Case and Number Marking

Hungarian case morphology occurred with a wide range of English-origin nominal material. Accusative marking via the suffix *-t* was found in forms such as *content-et*, *gameplay-t*, *spotlight-ot*, *groove-ot*, *cover-t*, and *progress-t*.

Directional and locative morphology included the illative *-ba/ -be*, the inessive *-ban/ -ben*, the superessive *-n*, the sublative *-ra/ -re*, the delative *-ról/ -ről* and the elative *-ból/ -ből* in forms such as *fullscreen-**be**, shop-**ba**, live-**ban**, stream-**en**, rooftop-**ra**, impro session-**re**, game-ről, fashion week-**ről*** and *mainstream-**ből***.

Other attested relations included instrumental *concurrent viewer-**rel***, *private jet-**tel***, and *small talk-**kal***; dative *reward-**nak*** and *bullshit-**nek***; causal-final *donate-**ért*** and *fake meat-ek-**ért***; essive *reset-**ként***; and adessive *rush-**nál***.

Number marking occurred in plural forms including *player-**ek***, *stream-**ek***, *scandal-**ök***, *outlet-**ek***, *bullshit-**ek***, *project-**ek***, and *setup-**ok***. Plural and case morphology frequently co-occurred, as shown in Table 14 below.

Table 14. Plural and case morphology applied on English-origin bases

	Examples	Glossing
Ex.29	*target-ek-re*	target.N-PL-SUBLAT
Ex.30	*highlight-ok-ban*	highlight.N-PL-INESS
Ex.31	*side quest-ek-et*	side.N quest.N-PL-ACC

Of the 168 lone English-origin nouns identified in the corpus, 91 (54.2%) occurred with overt Hungarian morphology, while 77 (45.8%) occurred without overt morphological marking. In all such cases, with both lone and multiword English-origin bases considered, plural morphology preceded case marking, following the Hungarian sequence:

English-origin base + plural + case

6.2.5 Possessive Morphology

Hungarian possessive morphology occurred in 13 mixed constructions. First-person singular forms included *gaming channel-**öm***, *jet lag-**em***, and *project-**em***. Second-person singular possession occurred in *split-**ed**-et* and *workshop-**od**-ra*. The corpus also contained first-person plural *challenge-**ünk*** and third-person plural *brand-**jük***.

Third-person singular marking occurred in *highlight-**já**-t*, *token-**je***, *tagline-**ja***, *legacy-**je***, and *workshop-**já**-n*. Possession could combine with additional morphology, as visible in *highlight-já-t* and *workshop-já-n*.

The form *feature-**e-i-t*** contained a particularly complex sequence as it contains possessive morphology, possessed-plural marking, and the accusative suffix in this order. Possessive integration was therefore attested both with individual English-origin bases and with multiword expressions such as *gaming channel-öm* and *jet lag-em*.

6.2.6 Coverbs

Altogether, six constructions combined a Hungarian coverb with an English-origin verbal base, and, at the same time, contained Hungarian verbalizing morphology (see Table 15 below):

Table 15. Hungarian coverbs combined with English-origin bases

	Examples	Glossing
Ex.32	*túl-hype-ol-t*	over.COV-hype.V-VBZ-PAST
Ex.33	*túl-hype-ol-va*	over.COV-hype.V-VBZ-ADVZ
Ex.34	*ki-dodge-ol-ni*	out.COV-dodge.V-VBZ-INF
Ex.35	*át-force-ol-ni*	over.COV-force.V-VBZ-INF
Ex.36	*meg ... cheat-el-ni*	COV.PERF... cheat.V-VBZ-INF
Ex.37	*el-spoiler-ez-ni*	away.COV-spoiler.N-VBZ-INF

Most coverbs immediately preceded the derived verb. In *meg akar-ja cheat-el-ni*, however, *meg* was separated from *cheat-el-ni* by the finite verb *akar-ja* ‘want.V-DEF’. The English-origin verb therefore participated in the discontinuous coverb configuration required by the Hungarian predicate structure.

6.2.7 Mixed Compounds

Mixed compounds and compound-like formations occurred in several configurations. Most frequently, the combination included an English-origin modifier and a Hungarian head, as in

review videó (video), win-win helyzet (situation), content gyártás (production), freestyle vers (poem), rant csatorna (channel), mental health téma (topic), viewing élmény (experience), donate gomb (button), believe felirat (label), peace jel (sign), and *trigger pont (point)*. The Hungarian head carried external inflection where required, as in *review videó-k-at* and *win-win helyzet-nek*.

The opposite order, with Hungarian material preceding an English-origin element, was less frequent but occurred in *gyakorlat (exercise) selection*, *teszt (test) content*, *sör (beer) -time*, and *téma (topic) -breaker*. Other formations included partially translated combinations such as *french krumpli (fries)* and *win-win szituáció (situation)*.

These forms were distinguished from entirely English sequences such as *feature set*, *cash grab*, *fake news*, *small talk*, and *content plan*. In individual occurrences, however, the boundary between compound, lexicalized expression, and productively generated phrase was not always structurally recoverable.

6.2.8 Multiword English Material and Potential EL Islands

The dataset contained 103 multiword constructions. Of these, 35 (34.0%) carried Hungarian morphology, while 68 (66.0%) contained no Hungarian intraword morphology.

The morphologically integrated group included examples like *sleep tracking-re, live stream-ek-et, community guideline set-tel, join request-et, french fries-nak, private jet-tel, jam session-ök-be, side quest-ek-et, small talk-kal,* and *fake meat-ek-ért*.

The non-intraword group included both short English expressions and longer internally structured sequences. Short forms included *feature set*, *cash grab*, *fake news*, *wheel bite*, *happy end*, and *drag queen show*; while internally structured forms included *forever and ever, nothing else*, *been there, done that*, *no education just liberation*, and *in house*.

Sequences of the latter type provided clearer evidence of internal English grammar because they contained function words, verbal forms, coordination, or other internal dependencies. Short expressions such as *fake meat*, *small talk*, and *feature set* were less structurally diagnostic because they could also function as lexical units or compounds, yet they are useful in testing the predictions of the frameworks too.

Thus, the multiword data consisted of structurally different configurations rather than a single homogeneous category. They contain internally English sequences without Hungarian

morphology, multiword bases receiving Hungarian external morphology, and mixed English-Hungarian constructions.

6.3 Recategorization

Recategorization was coded where Hungarian morphology assigned an English-origin base to a grammatical category different from its identifiable source-language category. This was distinguished from category-maintaining integration, since many English forms such as *stream*, *force*, *review*, *coach*, *flex*, *update*, and *hype* can already occur in more than one lexical category in English.

The clearest verbal recategorizations involved nominal or multiword bases converted into Hungarian verbs:

roadshow	→	*roadshow-z-ik*	(N-VBZ-3S.INDEF)
spoiler	→	*spoiler-ez-ni*	(N-VBZ-INF)
fingerflip	→	*fingerflip-el-ni*	(N-VBZ-INF)
new age	→	*new age-el-t-ek*	(N-VBZ-PAST-3PL)
face app	→	*face app-oz-ott*	(N-VBZ-PAST)

These forms illustrate the type of recipient-language categorization strategy discussed by Meechan and Poplack (1995), with the Hungarian verbalizer overtly establishing the category required by the construction.

Some forms underwent more than one categorial operation. In *gameplay-ez-és-nél*, the English nominal base was verbalized by *-ez*, and subsequently nominalized by *-és*, and finally marked for case. *Stand-up-ol-ás-t* displayed the same sequence:

English nominal base → Hungarian verb → Hungarian deverbal noun → inflected noun

Recategorization also occurred outside the verbal domain in examples where, for example, the multiword nominal base *pick-up truck* received adjectival *-os* in *pick-up truck-os*, while *happy* was converted into the abstract noun *happy-ség*. Forms such as *sales-es*, *stand-up-os*, and *late-night show-s* produced person-denoting or adjectival constructions from English-origin bases. Interestingly, the plural *feel good-ok* similarly treated the English expression *feel good* as a Hungarian countable nominal base.

Morphologically unmarked uses such as *hype*, *open*, and *full* were not treated as conclusive recategorization because English itself permits extensive conversion and categorial flexibility. On the other hand, overt Hungarian category-forming morphology provided quite clear evidence of it.

6.4 Testing the MLF Predictions

The MLF analysis tested the five predictions formulated in Section 4.5.1. In brief, they concern Hungarian Matrix Language status, Hungarian ordering at diagnostic conflict sites, Hungarian outsider late system morphology, conformity of English-origin lexical material to Hungarian grammatical requirements, and the treatment of internally English multiword material (EL islands).

6.4.1 Matrix Language Identification

Hungarian was identifiable as the Matrix Language in all 405 intraclausal constructions. ML identification was based on the combined clause-level evidence specified in Chapter 5: predicates, function words, argument structure, constituent organization, and grammatical morphology.

This pattern was particularly transparent where English-origin nominal material occurred within clauses containing Hungarian predicates and Hungarian argument marking, for example:

*'sajnál-om a **badge**-et'* (*I am sorry for the badge*)
feel.sorry.V-1S.DEF the.DET.DEF badge.N-ACC

*'azonnal nyit-nak egy **ticket**-et'* (*They open a ticket immediate*ly)
immediately.ADV open.V-3PL a.DET.INDEF ticket.N-ACC

*'te ez-t a **game**-et jobb-an élvez-t-ed?'* (*Did you enjoy this game more?*)
you.PRON.PERS this.PRON.DEM-ACC the.DET DEF game.N-ACC better.ADJ-ADVZ enjoy.V-PAST-2S.DEF

English-origin verbal material showed the same clause-level asymmetry. Forms such as *stream-el-ünk*, *force-ol-ják*, and *hijack-el-t-ék* occurred as Hungarian-derived predicates carrying Hungarian grammatical morphology.

Multiword English sequences also did not alter the ML identification of the containing clause. Forms such as *feature set*, *cash grab*, *mental health*, and *been there, done that* occurred within otherwise Hungarian-framed structures.

Conclusively, in no intraclausal construction was English identified as the language supplying the clause-level morphosyntactic frame, making this result consistent with the MLF prediction of structural asymmetry (Myers-Scotton, 1993, 2002; Deuchar, 2006).

6.4.2 Morpheme Order Principle

The Morpheme Order Principle was evaluated at points where Hungarian and English supplied distinguishable ordering predictions. The clearest evidence came from the 187 intraword constructions, all of which followed Hungarian morphological order.

Complex examples (presented and analysed earlier) included:

gameplay-ez-és-nél

feature-e-i-t

leg-light-os-abb-ak

fake meat-ek-ért

community guideline set-tel

side quest-ek-et

Similarly, the 40 verbal constructions followed the Hungarian sequence where the English-origin base was followed by a Hungarian verbalizer and any additional derivational or inflectional morphology. Again, no construction presented a different ordering.

The six coverb constructions provided syntactic evidence in addition to suffix ordering. *Ki-dodge-ol-ni*, *át-force-ol-ni*, *el-spoiler-ez-ni*, and *túl-hype-ol-t* followed Hungarian predicate organization, while *meg akar-ja cheat-el-ni* displayed appropriate Hungarian coverb placement.

Nevertheless, not all constructions were diagnostic, as modifier – head combinations such as *fake meat*, *mental health téma*, and *review videó* were compatible with ordering patterns available in both languages, and so, they were not considered as independent tests of the MOP, however, no unambiguous MOP violation was identified, which shows complete consistency among the possible diagnostic constructions.

6.4.3 System Morpheme Principle and 4-M Predictions

The System Morpheme Principle was evaluated through the source of outsider late system morphology in ordinary ML+EL constituents. In every case where grammatical morphology depended on relations external to the English-origin lexical head, the overt material was Hungarian. Structurally assigned case was Hungarian in forms such as *content-et*, *game-ről*, *reset-ként*, and *private jet-tel*. English-origin predicates received Hungarian tense, mood, person, number, and morphemes of definiteness, as in *stream-el-ünk*, *force-ol-ják*, *discount-ol-nám*, *hijack-el-t-ék*, and *jam-el-j-ünk*. Possessive agreement was likewise Hungarian in *brand-jük*, *gaming channel-öm*, *challenge-ünk*, *tagline-ja*, and *workshop-od-ra*.

Again, no English finite agreement, tense, structurally assigned case, or possessive agreement morphology occurred in an any Hungarian ML+EL constituent.

English grammatical morphology was restricted to internally English material such as *men's rights activism*, *no dead babies*, and *been there, done that*. However, these sequences were evaluated separately as potential EL islands.

While morphologically bare English-origin items did not contradict the SMP as the corresponding Hungarian structure did not require overt affixation, no clear English outsider late system morpheme was identified within any Hungarian ML+EL constituent, which is consistent with the predictions of the SMP.

The 4-M analysis distinguished the contribution of content, early system, bridge late system, and outsider late system morphemes (Myers-Scotton, 2002; Myers-Scotton & Jake, 2000), and predicts the latter to be sourced from the ML in all cases. This prediction was not violated, as English primarily provided inserted content morphemes, including nouns, verbal bases, and adjectives such as *content*, *badge*, *stream*, *clean*, and *toxic*, together with complex English-origin bases like *sleep tracking*, *trash talk*, and *fake meat*; whereas all clearly identifiable outsider late system morphology (e.g. structurally assigned case and agreement morphemes) in ordinary ML+EL constituents was Hungarian, for example in *brand-jük* 'brand.N-3PL.POS' and *workshop-od-ra* 'workshop.N-2S.POS-SUBLAT'.

6.4.4 Embedded Language Islands

The 103 multiword constructions were examined for independently identifiable English internal grammar. Such potential EL islands were:

nothing else
the end
forever and ever
no dead babies
been there, done that
no education, just liberation
so much fun

Since these sequences contained English function words, inflectional material, coordination, or other internal grammatical dependencies and retained English internal organization within Hungarian discourse, their English morphology and word order were compatible with EL-island status in the MLF Model.

Shorter English expressions such as *feature set*, *cash grab*, *fake news*, *small talk*, *drag queen show*, and *mental health* provided weaker evidence because their internal organization could reflect compounds or lexicalized expressions rather than productive English syntax, yet they are also allowed within the MLF.

6.4.5 Problematic Cases

The main analytical difficulty concerned the internal status of complex English material. The 35 morphologically integrated multiword switches could be represented within the MLF analysis as complex EL material receiving Hungarian external morphology, but the data did not always determine whether the English sequence itself was a compound, a lexical unit, or a productively structured subconstituent. Nonetheless, none of the examples constituted an unambiguous violation of the MOP or the relevant SMP/4-M prediction

Congruent ordering also reduced diagnostic strength, but it did not violate MLF predictions. Forms such as *fake meat* and *review videó* were compatible with Hungarian structure but did not distinguish Hungarian from English ordering independently.

Two expressions, *in the go* and *have a fun*, also consisted entirely of English material and displayed English internal constituent structure, although their forms differed from

conventional standard-English usage. While their non-standard realisation was noted, they were treated as potential EL island cases rather than as evidence against the MLF principles.

In conclusion, at the framework level, all 405 intraclausal constructions received an MLF-compatible structural analysis. Although the diagnostic strength of individual constructions varied, no applicable prediction was contradicted.

6.5 Testing the Poplackian Predictions

The Poplackian analysis tested the three predictions formulated in Section 4.5.2, which concern L^R morphosyntactic patterning of lone English-origin items, the possibility of L^D integration into the L^R without prior social establishment under the NBH, and retention of identifiable English internal grammar in multiword code-switches (Poplack, 2012, 2018; Poplack & Meechan, 1998; Sankoff et al., 1990).

6.5.1 Lone English-Origin Items

Lone English-origin material accounted for the majority (302) of the 405 intraclausal constructions (74.6%). Of which, 152 (50.3%) carried overt Hungarian morphology, while 150 (49.7%) occurred without morphology attached directly to the English-origin form. The overtly integrated group included forms such as *content-et, stream-el-ünk* and *pure-abb*; and provided the clearest token-level evidence of Hungarian L^R morphosyntax.

The 150 morphologically unmarked items were not automatically classified as unintegrated, since their syntactic placement was according to the Hungarian frame. Examples from Table 16 show that they occurred in environments where the corresponding Hungarian category could also appear without overt morphology.

Table 16. Lone English-origin items without overt morphology

Ex.38	Examples	Translation
Ex.39	*'ez is content'*	*this is content too*
Ex.40	*'ez teljesen fine'*	*this is totally fine*
Ex.41	*'van aki azt mondta, hogy cringe'*	*someone said that it is cringe*
Ex.42	*'akkor sorry'*	*sorry then*

Bare tokens were less diagnostic than overtly marked ones, particularly where the surface structure was compatible with both languages, although this treatment is consistent with Owens (2005), who cautions against equating absence of L^R morphology automatically with absence of integration.

6.5.2 Morphosyntactic Integration

As predicted by the framework, all 152 overtly integrated lone items followed Hungarian morphological structure. Nominal bases received Hungarian case, number, and possession, as illustrated by examples like *reset-ként*, *rush-nál*, *player-ek*, *project-em*, and *feature-e-i-t*, while English-origin verbal material received Hungarian verbalization and subsequent Hungarian inflection, including *stream-el-ünk*, *force-ol-ják*, *discount-ol-nám*, *spoiler-ez-z-em*, *update-el-ve*, and *flex-el-és-ről*. Adjectival and categorial morphology occurred as well in *full-os*, *sales-es*, *heavy-bb*, *leg-light-os-abb*, and *happy-ség*, further supporting the consistency of the prediction treating such forms as borrowings.

No overtly integrated lone item used English case, finite agreement, or other English grammatical morphology to realize a relation required by the surrounding Hungarian clause, so these findings were consistent with the Poplackian prediction that borrowing-compatible lone items follow L^R morphosyntax (Poplack & Meechan, 1998; Poplack, 2018).

As the present analysis establishes the Hungarian morphological behaviour of the attested forms, it does not, however, constitute a full variationist equivalence test across matched native Hungarian lexical items, since the corpus did not contain a separate benchmark dataset for quantitative comparison in equivalent variable contexts.

6.5.3 Established and Nonce Forms

Dictionary-attested established Hungarian loanwords were excluded during the initial identification procedure. The retained dataset therefore did not contain forms treated as established loans in the Hungarian reference source used for filtering.

Non-attestation did not establish nonce status. Forms such as *content*, *mainstream*, *cool*, *social media*, and *cancel culture* may be familiar or recurrent beyond the present corpus despite

their absence from the dictionary criterion applied here. Conversely, a low number of occurrences within the corpus was insufficient to demonstrate social novelty.

Nevertheless, several apparently low-recurrence forms showed immediate Hungarian grammatical integration. Examples such as *spotlight-ot* and *fingerflip-el-ni* received Hungarian morphology in their attested occurrences and consequently were compatible with the NBH prediction that grammatical integration does not depend on prior establishment (Sankoff et al., 1990; Poplack, 2012).

The corpus provided NBH-compatible evidence, but not a categorical established-versus-nonce classification of the English-origin items. Independent community-level evidence of establishment was unavailable, so low frequency and dictionary non-attestation were not used as sufficient criteria for nonce borrowing.

6.5.4 Multiword Switches

Of the 103 multiword constructions, 68 (66.0%) contained no Hungarian intraword morphology and fell within the canonical multiword CS domain of the Poplackian comparison. The clearest cases providing evidence for the prediction concerning CS include *nothing else, let's go, so much fun* and *no education, just liberation*.

As these sequences contained English internal dependencies, they retained L^D morphosyntax within Hungarian discourse. Here, the short multiword expressions such as *feature set*, *cash grab*, *fake news*, *happy end*, and *drag queen show* were also less diagnostic, as their internal structure could also correspond to compounds, fixed expressions, or complex lexical units rather than productively generated phrasal switches.

Mixed formations such as *review videó*, *content gyártás*, and *mental health téma* likewise did not provide the same type of L^D-internal evidence as the longer English sequences.

Thus, the strongest support for the multiword code-switching prediction came from constructions displaying independently identifiable English grammar, while shorter expressions remained compatible with the framework's analysis but supplied weaker evidence.

6.5.5 Problematic and Indeterminate Cases

The 35 multiword constructions carrying Hungarian intraword morphology (see Appendix D) were proven to be indeterminate as the framework does not provide sufficient diagnostic to

describe and analyse these forms. Such examples represented 34.0% of the multiword constructions and 8.6% of the complete intraclausal dataset and included constructions from Table 11 and the ones below:

gaming channel-öm

live stream-ek-et

side quest-ek-et

new age-el-t-ek

All 35 displayed overt Hungarian morphological integration, but the base receiving the morphology contained more than one English lexical element, thus lacking requirements for a borrowing classification.

These constructions did not correspond to the canonical contrast between a L^R-integrated lone item and a multiword code-switch retaining L^D grammar. At the same time, their Hungarian morphology did not constitute a counterexample to the Poplackian approach, as borrowing may involve compounds and complex lexical units (Sankoff et al., 1990), and multiword expressions may become available as lexical units in bilingual speech (Backus, 2003).

The individual tokens generally did not provide sufficient evidence to determine which analysis can be applied. In *fake meat-ek-ért*, for example, Hungarian plural and case morphology clearly establish external Hungarian integration, but the internal adjective – noun sequence *fake meat* does not distinguish productive English structure from a complex lexical unit because the same ordering is compatible with Hungarian. A similar ambiguity occurred in *sleep tracking-re*, *small talk-kal*, and several other forms.

The 35 constructions were therefore classified as indeterminate under the Poplackian analysis. This classification concerns their borrowing/code-switching status, not the observable fact of Hungarian morphological integration. Meanwhile, bare lone items and short multiword expressions could also provide weaker diagnostic evidence than overtly integrated or internally structured forms, but they were not added to the framework-level indeterminate category as they do not violate the predictions.

6.6 Comparative Framework Coverage

Following the framework-specific analyses in Sections 6.4 and 6.5, the overall analytical outcomes were compared across the 405 intraclausal constructions. As defined in Chapter 5, constructions were classified at framework level as consistent, inconsistent, or indeterminate. Table 17 summarises the coverage of the two frameworks on all intraclausal cases.

Table 17. Compared framework coverage of the data

Framework	Consistent cases	Inconsistent cases	Indeterminate cases	Total
MLF	405 (100.0%)	0 (0.0%)	0 (0.0%)	405
Poplackian	370 (91.4%)	0 (0.0%)	35 (8.6%)	405

Conclusively, under the MLF analysis, all 405 intraclausal constructions received a determinate structural analysis. Hungarian was identified as the Matrix Language throughout the dataset, and no construction showed an unambiguous violation of the applicable predictions.

Within the Poplackian framework, 370 constructions received a determinate analysis. These comprised the 302 lone English-origin items and the 68 multiword constructions without Hungarian intraword morphology, while the remaining 35 constructions combined multiword English material with Hungarian morphological integration and were classified as indeterminate. No construction was classified as a clear contradiction of either the MLF or the Poplackian predictions. Therefore, the main difference between the framework-level outcomes was the 35 indeterminate Poplackian cases, representing 8.6% of the intraclausal dataset. The theoretical significance of these results is considered in the following chapter.

7. DISCUSSION

7.1 Summary of the Major Findings

The findings show extensive and systematic integration of English-origin material into Hungarian morphosyntax. Of the 419 English-origin occurrences identified in the corpus, 405 were intraclausal and formed the basis of the structural comparison. These included 302 lone English-origin items and 103 multiword constructions. Intraword integration occurred in 187 cases, comprising 152 lone items and 35 multiword constructions, where English-origin material participated in Hungarian case, number and possessive marking, verbal and adjectival derivation, degree formation, nominalization, verbal inflection, and coverb constructions. Several forms contained multiple successive layers of Hungarian derivational and inflectional morphology.

This structural pattern answers the first research question clearly: English-origin material was not merely inserted alongside Hungarian material but regularly entered grammatical structures organized by Hungarian. Particularly revealing examples included verbal formations such as *stream-el-ünk*, multilayered forms such as *gameplay-ez-és-nél* and *feature-e-i-t*, and externally inflected multiword expressions such as *fake meat-ek-ért* and *community guideline set-tel*; however, constituent size and morphological integration did not vary categorically together, as Hungarian morphology could apply to both lone lexical items and to complex English-origin bases.

The MLF analysis provided a determinate structural account of all 405 intraclausal constructions, with Hungarian being identifiable as the Matrix Language throughout the corpus. Diagnostic intraword constructions followed Hungarian morpheme order, and the relevant outsider late system morphemes came from Hungarian. Also, English-origin lexical material conformed to Hungarian category-forming and inflectional requirements, and internally English sequences could be integrated through the distinction between ML+EL constituents and Embedded Language islands. In sum, no unambiguous violation of the applicable of the model's predictions was identified.

The Poplackian framework also accounted for most structural tendencies, with 152 overtly integrated lone items displaying Hungarian morphosyntax, while the clearest internally structured multiword sequences retained English grammar. These findings were compatible

with the distinction between L^R-integrated borrowing and L^D-structured multiword code-switching.

Evidence for the NBH was more limited: apparently non-established forms could display immediate Hungarian integration, which is predicted by the model, but their nonce status could not normally be established independently. What is more, 35 morphologically integrated multiword constructions remained indeterminate under the canonical Poplackian analysis.

The most important comparative result is probably that the grammatical structure of a bilingual construction can be considerably clearer than its classification as CS or borrowing. While both frameworks recognize the Hungarian contribution to the external structure of the data, they do not assign the same theoretical status to the English-origin material. This distinction between structural description and contact-linguistic classification is discussed in the following sections.

7.2 Empirical Evaluation of the MLF Model

The findings of the study provide strong support for the structural asymmetry predicted by the MLF Model, as Hungarian consistently supplied the clause-level morphosyntactic frame, while English contributed primarily lexical content and, in some cases, longer internally structured constituents. This distribution corresponds closely to the organization of classic code-switching proposed by Myers-Scotton (1993, 2002), in which the participating languages make unequal contributions to the bilingual constituent.

Hungarian Matrix Language status was supported by converging morphosyntactic evidence and not only by the numerical predominance of Hungarian words. It provided predicates, function words, argument structure, constituent organization, and grammatical morphology. Especially transparent distinction could be seen in constructions such as *content-et*, *brand-jük*, and *stream-el-ünk*, where the English-origin component provided lexical content while Hungarian determined its grammatical realisation.

The verbal domain was informative as well, since all 40 English-origin verbal constructions contained Hungarian verbalizing morphology before any further derivation or inflection. Forms such as *stream-el-ünk*, *force-ol-ják*, *spoiler-ez-z-em*, and *discount-ol-nám* reproduce the structural pattern identified by Bolonyai (2005), where Hungarian verbalizers overtly establish the categorial status required for participation in the Hungarian verbal system. The dissertation's findings extend this pattern to a broader, naturalistic corpus and, importantly, to

complex bases such as *trash talk-ol-t* and *new age-el-t-ek*. Thus, the grammatical frame can accommodate EL material with lexical complexity that exceeds a single orthographic word.

As all 187 intraword constructions followed Hungarian morphological ordering, the prediction of the Morpheme Order Principle was clearly supported in structures where Hungarian and English have distinguishable grammatical needs. Forms containing several successive morphemes, such as *gameplay-ez-és-nél*, *feature-e-i-t*, and *fake meat-ek-ért* were particularly informative. Bilingual verb – coverb constructions added further evidence: *meg akar-ja cheat-el-ni*, for example, followed the Hungarian syntactic placement of the coverb despite the English origin of the verbal base.

Many constructions were compatible with both Hungarian and English ordering and were treated as non-diagnostic, so absence of MOP violations must be distinguished from the strength of the evidence. Modifier – head sequences such as *fake meat* could not establish a ML word order independently where the languages permitted the same arrangement. As the results are giving the strongest evidence at actual conflict sites, they do not necessarily provide 405 equally informative confirmations of the MOP.

A similar caution is needed for the analysis of the System Morpheme Principle. The most direct evidence came from outsider late system morphemes. All clearly identifiable outsider late system morphemes in normal ML+EL constituents were Hungarian. English-origin predicates received Hungarian person, number, tense, mood, and definiteness morphology, while structurally assigned case and possessive agreement were likewise Hungarian. English outsider late system morpheme was not identified in any Hungarian-framed ML+EL constituent. This result is consistent with the later 4-M-modified SMP's prediction (Myers-Scotton, 2002; Myers-Scotton & Jake, 2000).

By contrast, Hungarian verbalizers, derivational morphology, and many other grammatical elements were early system morphemes. Their Hungarian source fits the broader structural asymmetry of the corpus and is allowed concerning the MLF, but it should not be counted as equivalent evidence for the SMP as late outsiders.

The EL-island mechanism also proved useful and provided additional evidence for the success of the MLF model. Clearly internally structured English sequences such as, *nothing else*, and *been there, done that* retained English grammar without contradicting Hungarian ML control. The distinction between an English-internal constituent and the larger Hungarian frame

captured a pattern that cannot be represented well by treating every occurrence of English morphology as evidence against Hungarian structural dominance.

The 35 morphologically integrated multiword constructions are especially significant in this respect (see Appendix D). Forms such as *fake meat-ek-ért*, *community guideline set-tel*, and *small talk-kal* could be analysed as complex English-origin material whose external grammatical realization was supplied by Hungarian. Because the MLF Model does not require EL content material to correspond to a single orthographic word, it can accommodate the coexistence of internal English lexical complexity and external Hungarian morphology.

Nonetheless, the complete analytical coverage of the MLF Model should be interpreted with care. ML identification must remain sufficiently independent of the further predictions tested, because if the same Hungarian morphology were used both to identify the Matrix Language and to confirm that Matrix Language morphology is Hungarian, the reasoning would become circular. In the present study, this problem was reduced by identifying Hungarian ML status from a wider combination of morphosyntactic evidence before evaluating MOP and SMP predictions separately.

The flexibility of MLF categories such as ML+EL constituents, EL islands, structural congruence, and accommodation allow the model to represent a wide range of bilingual configurations. This is one of its strengths, but it also means that categories must be justified independently.

In essence, the MLF Model performed well on the corpus. This did not come from the fact that no counterexample was found, but that it supplied a coherent structural representation across lone items, complex intraword forms, English-derived predicates, coverb constructions, multiword bases, and internally English multiword constituents.

7.3 Empirical Evaluation of the Poplackian Approach

The Poplackian analysis was also well supported in its clearest empirical domains. Lone English-origin items frequently behaved according to Hungarian morphosyntax, while internally structured longer English sequences retained identifiable English grammar. This reproduces the central contrast between L^R-integrated borrowing and L^D-structured multiword code-switching developed in the variationist tradition (Poplack, 2018).

The 152 overtly integrated lone items provided the strongest evidence for the framework's success. English-origin nouns received Hungarian case, number, and possessive morphology and English bases underwent Hungarian verbal and adjectival derivation, which was followed by additional grammatical marking where required. So, these forms exhibited precisely the kind of L^R behaviour that is relevant to a borrowing analysis.

Of these lone integrated items, the verbal constructions again provided particularly clear evidence. English-origin bases entered the Hungarian verbal paradigm through Hungarian verbalization and could subsequently receive Hungarian morphology. Their English lexical origin did not result in the retention of English verbal grammar within the surrounding Hungarian construction, which from the Poplackian view, is strong evidence of L^R integration.

The corpus also contained forms where the infrequent occurrence was compatible with the Nonce Borrowing Hypothesis. Apparently novel items such as *fingerflip-el-ni* received Hungarian grammatical morphology immediately in the recorded occurrence, which is in accordance with the NBH claim that an item does not need to become socially established before acquiring L^R grammatical characteristics (Sankoff et al., 1990; Poplack, 2012).

However, the findings do not, constitute a complete test of the NBH. The strongest variationist demonstrations compare L^D-origin items with native L^R material under equivalent grammatical conditions. The relevant issue is not only whether an English-origin noun can receive a Hungarian case suffix, but whether its marking follows the same grammatical conditioning as comparable Hungarian nouns (Poplack & Meechan, 1998; Poplack, 2018). The present corpus establishes extensive Hungarian integration, but it was not designed to match variable – rule comparison with native Hungarian lexical items.

This is illustrated by the 150 morphologically unmarked lone items, where in many cases, zero marking was fully compatible with the corresponding Hungarian environment. Therefore, such forms cannot simply be labelled unintegrated. At the same time, a token compatible with both Hungarian and English provides weaker evidence than a construction in which the languages impose different grammatical requirements. Still, syntactic integration occurred in all cases, which is predicted by the prediction of the NBH. Owens (2005) as well cautions against interpreting this absence of overt L^R morphology as categorical evidence of non-integration.

Establishment presented another evidential limitation. Dictionary-attested conventional loans had been removed during initial corpus construction, but non-attestation in the selected dictionary did not establish that the retained forms were genuinely nonce. Expressions such as *content*, *mainstream*, or *social media* may (and probably do) already circulate widely within certain Hungarian communities despite their absence from a general dictionary, nor does a single occurrence in the corpus establish novelty. Consequently, the corpus supports immediate grammatical integration better than claims about the social status of individual lexical items.

This limitation is especially relevant to the debate surrounding the NBH. As discussed in previous chapters, Stammers and Deuchar (2012) found frequency-related differences in the application of Welsh soft mutation to English-origin verbs and interpreted them as problematic for the hypothesis. Poplack (2012), however, argued that the NBH does not predict a developmental sequence in which grammatical integration gradually increases with frequency; it predicts that L^R grammatical behaviour can be available from the outset. The present findings are compatible with this weaker but central claim, while lacking the community-level evidence required to determine the nonce status of individual items independently.

Multiword results also supported the Poplackian prediction where English internal grammar was independently identifiable. Expressions such as *let's go*, and *been there, done that* retained English grammatical organization. Shorter expressions such as *feature set*, *fake news*, *small talk*, and *drag queen show* were less diagnostic because they could also function as compounds or lexicalized units but nonetheless can be interpreted as multiword CS. Thus, multiword status itself was insufficient; the relevant evidence was identifiable L^D internal structure.

The most substantial issue involved the 35 constructions combining multiword English material with Hungarian morphology. Forms such as *fake meat-ek-ért*, *sleep tracking-re*, *gaming channel-öm*, and *trash talk-ol-t* clearly showed Hungarian external integration but contained more than one English lexical component. They could not be classified confidently through the canonical opposition between integrated lone items and L^D-structured multiword code-switches.

This does not make them direct counterexamples to the Poplackian approach, as Sankoff et al. (1990) do not restrict borrowing categorically to single orthographic words, and according to them, complex lexical material may enter another language. Backus (2003) similarly shows that compounds, collocations, idioms, and other multimorphemic sequences may function as lexical units in bilingual speech. The difficulty is evidential: a single naturalistic occurrence

often does not reveal whether a sequence such as *fake meat* is being compositionally generated as English syntax, accessed as a lexical unit, or integrated through some intermediate representation.

All considered, the Poplackian analysis achieved broad but less successful analytical coverage than the MLF Model in this corpus. Its strongest contribution was the requirement that claims about borrowing be grounded in grammatical behaviour rather than etymology, pronunciation, frequency, or dictionary status alone, while its main limitations arose from the evidence available: incomplete community-level information, the absence of matched Hungarian variationist benchmarks, and the indeterminate status of morphologically integrated multiword expressions.

7.4 Where the Frameworks Converge

The comparison reveals considerable agreement before the frameworks diverge in interpretation. Most importantly, both approaches treat the bilingual constructions as grammatically systematic. The presence of English-origin material does not suspend Hungarian grammatical organization, nor is switching treated as an arbitrary juxtaposition of elements from two languages.

Both approaches also recognize that lexical origin and grammatical behaviour are separate properties. An English-origin lexical item can participate in Hungarian morphosyntax, while a larger English constituent may preserve English internal organization. This means that the language from which a lexical form originates is not a sufficient diagnostic to determine how it functions in the bilingual construction.

The clearest convergence concerns the source of external grammatical structure. In forms such as *content-et*, *brand-jük*, and *stream-el-ünk*, both frameworks identify the relevant morphology as Hungarian and recognize that the English-origin material participates in Hungarian grammatical relations. In the MLF, this is described as ML control over EL material, whereas under the Poplackian approach, it is L^R integration. The theoretical labels differ, but their structural observation is shared.

A second point of convergence concerns internally English material. Longer sequences containing clear English grammatical dependencies retain their English internal organization, which, according to the MLF, is through EL islands, while the Poplackian approach treats such

L^D-structured stretches as code-switches. Again, the frameworks reach compatible structural descriptions even though they arrive here through different theoretical concepts.

Finally, neither framework requires overt Hungarian affixation on every English-origin item. A bare item may be compatible with Hungarian grammar where the relevant Hungarian category is itself unmarked, and this is quite important because treating overt suffixation as the sole measure of integration would misrepresent both approaches.

Although the descriptive convergence of the frameworks is substantial, their main disagreements do not arise over the identification of Hungarian and English material, but over the theoretical consequences assigned to the resulting grammatical configurations.

7.5 Where the Frameworks Diverge

As stated before, the MLF Model explains the internal architecture of bilingual clauses and constituents, with its central question concerning which language supplies the morphosyntactic frame and how material from the other language is accommodated within it. Meanwhile, the Poplackian approach is mostly concerned with identifying the language-contact process responsible for L^D-origin material, especially the distinction between borrowing and CS. So, the frameworks overlap empirically without being equivalent theoretical alternatives. Their most fundamental divergence is based on, in short, how they interpret the same morphosyntactic evidence. They are applied here to the same empirical phenomenon and the same set of bilingual constructions, but they introduce different mechanisms, categories, and evidential requirements in order to determine what those constructions represent.

This is clearest with lone English-origin items. According to the MLF Model, an English content morpheme may occur inside a Hungarian ML frame and receive Hungarian morphology without ceasing to be EL material. On the other hand, the Poplackian approach, treats L^R-equivalent grammatical behaviour as evidence for borrowing. This way, a form such as *brand-jük* can satisfy both sets of structural expectations while receiving a different contact-linguistic classification.

Therefore, the roles of Matrix Language and recipient language must remain conceptually distinct even when Hungarian occupies both positions in a particular example. The ML is a structural relation within the MLF account of the bilingual constituent, while the L^R belongs to the analysis of how L^D-origin material is incorporated into it. The fact that both labels identify

Hungarian in this corpus does not make the concepts interchangeable, however they are suitable to carry out independent corpus analyses.

The frameworks also differ in the evidence that is needed to reach a determinate analysis. Several MLF predictions can be evaluated locally within an individual word or clause: morpheme order, system-morpheme source, category formation, and internal constituent structure are directly observable. At the same time, the Poplackian classification of borrowing, specifically the distinction between established and nonce borrowing, may additionally require comparison with native L^R forms and evidence concerning recurrence, recognition, and social distribution. This difference affects the strength of the evidence because a construction may contain enough structural information for an MLF analysis while lacking enough information on social distribution for a categorical borrowing analysis.

Morphologically integrated multiword constructions make this difference especially noticeable. The MLF can represent them as complex English-origin material within a Hungarian ML+EL construction, whereas the Poplackian analysis can identify their Hungarian integration but cannot always determine whether it is a complex borrowing, a lexical unit, a compound, or a productively structured code-switch. These examples account for the major difference between the frameworks; the MLF provided a determinate analysis for all 405 constructions, while 35 remained indeterminate under the analysis of the Poplackian framework.

This difference in coverage is meaningful, but it should not be described as a difference in statistical "accuracy." The percentages measure how much of the same dataset each model could classify determinately according to the operational criteria established in Chapter 5. The MLF had broader analytical coverage on the present corpus, while the Poplackian analysis imposed evidential conditions that could not be met for all constructions. These facts alone do not establish a theoretical superiority; however, they do indicate a different success rate of dealing with morphosyntactic integration in a typologically distant language pair.

Important to note that there is also a corresponding trade-off of the models' adaptability rates. The flexibility of MLF allows it to represent structurally diverse bilingual configurations, but this flexibility makes independent category definition essential. The stricter evidential requirements of the Poplackian approach reduce the risk of classifying items from surface foreignness alone, but this rigidity can leave structurally important naturalistic cases unresolved. The comparison therefore concerns not only whether a prediction is violated or not,

but also how much of the empirical domain each framework can analyse determinately and how much evidence it requires to do so.

To provide a clear visual of how the framework-specific predictions were evaluated, Table 18 below summarises their main outcomes.

Table 18. Summary of the predictions and their outcomes

Framework / prediction	Main result	Outcome
MLF: Matrix Language	Hungarian ML in all 405 intraclausal constructions	**Consistent**
MLF: Morpheme Order	Hungarian order in all diagnostic cases	**Consistent**
MLF: System Morphemes	Outsider late system morphemes were Hungarian	**Consistent**
MLF: English-origin material	English bases followed Hungarian grammatical requirements	**Consistent**
MLF: EL islands	Clear English multiword sequences retained English internal grammar	**Consistent**
Poplackian: Lone items	Lone English-origin items generally followed Hungarian morphosyntax	**Consistent**
NBH	Low-recurrence items could show Hungarian integration	**Compatible, but not fully testable**
Poplackian: Multiword switches	Clear multiword switches retained English internal grammar	**Consistent**
Integrated multiword material	35 cases remained unresolved under the Poplackian analysis	**Indeterminate**

Overall MLF coverage	405/405 constructions	**100% determinate**
Overall Poplackian coverage	370/405 constructions	**91.4% determinate**

7.6 Morphological Integration without Theoretical Agreement

The most significant theoretical implication of the study is probably that morphosyntactic integration can be established empirically without simultaneously resolving the contact status of the integrated material. In most cases, the descriptive analysis is not controversial, as an English-origin lexical base is generally identifiable, onto which Hungarian grammatical morphology is attached, and the resulting form participates in Hungarian constituent structure. The disagreement arises only when this observation is converted into a model-specific category, in short, what that integration should be called.

Considering the form *content-et* 'content-ACC' at the descriptive level, the English-origin base *content* carries the Hungarian accusative suffix *-et*. This form exhibits Hungarian nominal morphology and occupies a Hungarian syntactic position. Under the MLF Model, *content* can be analysed as an EL content morpheme within a Hungarian Matrix Language grammar. The Hungarian case suffix is expected because this language provides the relevant morphosyntactic structure. The same example observed under the Poplackian approach, is evidence that the lone English-origin item behaves as part of Hungarian recipient-language grammar and is consequently borrowing-compatible. Therefore, neither analysis disputes what the suffix is or what grammatical relation it expresses.

The same contrast applies to more complex forms: *stream-el-ünk* contains an English-origin base, Hungarian verbalization, and Hungarian first-person plural morphology and *brand-jük* contains Hungarian possessive marking on an English-origin noun. The structural contribution of Hungarian is clear in both cases, but the disagreement is about whether this constitutes grammatical accommodation of EL material during code-switching or the grammatical integration of borrowed material.

Distinguishing the two matters because accommodation/integration is an observable property, whereas borrowing and code-switching are theoretical classifications. Treating properties and classifications as synonymous would prejudge the comparison. While the MLF

Model explicitly predicts extensive Hungarian morphology on English material in a code-switching analysis, the Poplackian approach uses comparable L^R behaviour as evidence for borrowing. Therefore, the very same evidence has different inferential force under the two frameworks.

This issue becomes even more apparent with the 35 morphologically integrated multiword constructions, as in *fake meat-ek-ért*, where Hungarian plural and case morphology attach to a base containing two English lexical items. Here, the external Hungarian structure is unambiguous, while the internal status of *fake meat* is less so: it may function as an English compound, a stored multiword lexical unit, or a compositionally produced English expression. The MLF can represent this externally integrated construction without resolving the lexical question first, but the Poplackian classification requires more evidence before deciding between borrowing and multiword switching, even though none of the two classifications would straightforwardly predict the occurrence of such an example.

Treffers-Daller's (2005) treatment of mixed compounds and inserted nominal groups provides a useful parallel here. According to her, constituent complexity and insertion are not mutually exclusive, and reinforcing this claim, the present data show that constituent size and morphosyntactic integration are independent dimensions. Lone items may be overtly marked or morphologically bare, while multiword material may either retain English internal organization without Hungarian morphology or function as the base for Hungarian morphological processes.

A complementary psycholinguistic perspective is provided by Johns and Dussias (2022), who compared the online comprehension of single-word insertions and multiword alternations in highly proficient Spanish-English bilinguals. Their pupillometric results showed no processing difference between the two types of language mixing and led them to argue that morphosyntactic integration is not necessarily the defining processing property of single-word insertion. Although their study concerns comprehension rather than the grammatical classification examined here, it provides independent support for treating constituent size and morphosyntactic integration as distinct dimensions without assuming that lone and multiword material necessarily represent categorically different bilingual processes.

This suggests that the 35 multiword morphologically integrated forms are more than inconvenient exceptions. They demonstrate that a simple structural opposition between integrated lone items and structurally independent multiword switches does not exhaust the

possibilities attested in naturalistic bilingual speech. Multiword lexicalization, compounding, category formation, and external inflection can intersect.

This supports a layered analytical procedure in which four questions are kept separate:

1. What is the lexical origin of the material?
2. Which language supplies its internal and external grammatical organization?
3. What contact-linguistic classification is justified by the available evidence?
4. What evidence exists for social establishment or nonce status?

There is not a single diagnostic which can answer all four questions, since English lexical origin does not necessarily prove code-switching, the presence of Hungarian morphology does not prove nonce borrowing, multiword status is not a direct proof of productive English syntax, and finally, dictionary non-attestation alone cannot prove social novelty.

The theory-neutral descriptive stage adopted in this dissertation is therefore theoretically consequential not just methodological. Referring to *English-origin material*, *intraword integration*, or *multiword English material* before assigning framework-specific categories prevents the outcome from being built into the dataset itself and allows later classification for further framework-specific research.

Comparing them also shows a degree of complementarity between the frameworks without eliminating their competition. The MLF provides a detailed representation of how the bilingual construction is assembled, and the Poplackian approach demands stronger evidence before grammatical integration is translated into a borrowing classification. Yet the resulting categories can still conflict, because if Embedded Language code-switching and nonce borrowing are treated as mutually exclusive analyses of the same occurrence, the two frameworks simply cannot be combined to make their disagreement disappear.

In summary, the important finding is that structural certainty and classificatory certainty do not need to coincide. Hungarian morphology frequently makes it possible to identify how an English-origin element is grammatically incorporated with considerable precision, while its status as a code-switch or borrowing remains dependent on the theoretical and distributional evidence.

7.7 Hungarian Morphology as a Test Case for Code-Switching Theory

Hungarian is theoretically valuable in this comparison because grammatical relations that may remain covert in other languages are frequently expressed through overt, ordered morphology. As a predominantly suffixing, morphologically rich language, it makes it possible to trace the contribution of the two languages not only across constituents but within individual morphological words.

This advantage is best illustrated by the 187 intraword constructions, where English-origin material received Hungarian case, number, possession, verbalization, adjectival derivation, comparison, nominalization, and verbal inflection. In complex forms such as *feature-e-i-t* or *gameplay-ez-és-nél*, several distinct grammatical operations can be identified within a single bilingual construction.

As all 40 English-origin verbal constructions contained Hungarian verbalizing morphology before further inflection, the pattern seen in the results confirms Bolonyai's (2005) analysis of Hungarian verbalizers as markers of categorial identity, while showing that the process extends productively across contemporary naturalistic data and even to multiword bases. Hungarian separates lexical source from grammatical category overtly because an English-origin expression may enter the construction as of any word class, Hungarian morphology will determine how it functions as a verb.

The same property makes recategorization clearly observable. Meechan and Poplack (1995) showed that bilingual categorial mismatches may require L^R strategies of category assignment, and in the Hungarian corpus, this extended across verbal, adjectival, person-denoting, and abstract-nominal formations. *Roadshow-z-ik*, *pick-up truck-os*, and *happy-ség* show that integration can involve the creation of a new, Hungarian grammatical category instead of simply applying inflections to an unchanged English lexical item.

The relevance of lexical category is also supported by recent cross-linguistic corpus evidence. Chi and Bell (2024), analysing Spanish-English and Mandarin-English data, found a significant association between part of speech and the likelihood of code-switching, with the relationship strongest in the immediate vicinity of switching points. Although their study does not address morphosyntactic integration or test either framework considered here, it still reinforces the value of examining bilingual material by grammatical category without treating switched lexical material as a structurally uniform class.

Hungarian morphology also provides strong testing environments for the MOP and 4-M-informed SMP. Ordered derivational and inflectional sequences permit ML morpheme order to be observed directly within the word, while overt agreement, structurally assigned case, and possessive morphology provide potential outsider late system morphemes whose language source can be evaluated. At the same time, the corpus shows why these diagnostics must remain distinct, since a Hungarian verbalizer and a Hungarian agreement suffix do not carry identical theoretical weight simply because they are both grammatical morphemes.

Coverbs add a further dimension by linking word formation to clause-level syntax. The position of *meg* in *meg akar-ja cheat-el-ni* shows that an English-origin verbal base can participate in a discontinuous Hungarian predicate construction. This example shows that integration is not limited to suffix attachment but extends into the syntactic organization of the predicate.

The integrated multiword forms are perhaps the most revealing consequence of Hungarian morphology. Expressions such as *sleep tracking-re* and *community guideline set-tel* show that Hungarian grammatical morphology can take structurally complex English-origin material as its base. They provide direct evidence of constituent size not predicting the degree of external morphosyntactic integration.

Similarly to how Poplack et al. (1989) used their Finnish-English data, Hungarian as another morphologically rich Uralic language can be used to investigate the grammatical behaviour of English-origin material and distinguish borrowing from code-switching. Nevertheless, it should not be presented as the one providing automatic answers. Overt morphology is absent in many grammatical environments. Also, Hungarian and English are congruent in some ordering relations, and the presence of a suffix alone cannot establish its 4-M category or prove L^R-equivalent variation. What Hungarian offers seamlessly is structural visibility: it reduces uncertainty about many of the grammatical facts while leaving their theoretical interpretation open.

This is why morphologically rich languages are important for code-switching theory. Their value lies in extending research to under-described language pairs and in creating diagnostic environments where assumptions about category formation, morpheme order, system morphology, lexical insertion, and borrowing can be tested more explicitly.

7.8 Morphophonological Integration as Supporting Evidence

Morphophonological integration provides an additional perspective on the morphosyntactic findings. Hungarian suffixation often requires phonologically conditioned allomorph selection, particularly through vowel harmony, as presented by the Hungarian case marker or number morphology, for example. When an English-origin base receives such morphemes, the speaker must determine both the grammatical suffix and its phonological realization.

Hayes et al. (2009) demonstrate that Hungarian speakers possess productive and probabilistic knowledge of vowel harmony that can be extended to novel forms too. This suggests that vowel-harmonic treatment of English-origin material does not need to imply prior lexical establishment. As presented in the data, relatively novel or occasional material was accommodated to Hungarian morphophonology immediately and productively.

However, Hungarian vowel harmony should, not be reduced to mechanical suffix selection based on the preceding vowel. As Rebrus and Törkenczy (2021) show, front/back harmony interacts systematically with morphological structure and paradigm uniformity. The harmonic properties associated with a root are inherited by morphologically complex forms and may override expected phonological restrictions. The relevance of this to the present analysis comes from harmonic suffix selection on English-origin bases which should be understood within the broader Hungarian morphology – phonology system not only as simple phonetic accommodation to the final segment of the English base. Although Rebrus and Törkenczy (2021) do not examine bilingual material, so their findings provide a description of the Hungarian system into which the English-origin forms are incorporated, not bilingual constructions.

An earlier analysis of a previous version of the corpus by Salamon and Szentgyörgyi (2026) examined 181 suffixed intraword tokens. This dataset preceded the expanded dissertation corpus and need to be distinguished from the present set of 187 intraword constructions; however, of the earlier 181 tokens, 133 contained inflectional morphology and 48 contained at least one Hungarian derivational suffix. The phonological classification was conducted by type. Among 172 distinct forms, full phonological integration accounted for 152 types (88.4%), partial integration for 16 (9.3%), compound adaptation for three examples (1.7%), and clear alphabetic pronunciation for one (0.6%).

Forms such as *brand-jük* illustrate extensive Hungarian-oriented phonological accommodation alongside Hungarian morphology. *Content-et* represents a more complex case in which the realization of the English-origin base and the selected accusative allomorph combine into a form compatible with Hungarian harmonic organization. Complex bases such as *fullscreen-be* further show that Hungarian suffix selection can operate at the external edge of multiword or internally complex English material.

The predominance of Hungarian-oriented adaptation is in accordance with Stefanich et al. (2019), whose cross-linguistic analysis of mixed words suggests that derivational and inflectional morphology can exert a strong influence on the phonological realization of mixed constructions. In the present data, morphophonological evidence generally points towards the same direction: English supplies the lexical base, while Hungarian contributes morphophonological organization.

The evidential value of this pattern must nevertheless remain secondary. Stefanich et al. (2019) emphasize that strong conclusions concerning bilingual intraword phonology require acoustic analysis and comparison with relevant monolingual productions. The present auditory classifications show a clear tendency but cannot establish conclusively whether certain forms instantiate one phonological system, two interacting systems, or intermediate bilingual realisations.

Most importantly, phonological integration cannot independently distinguish code-switching from borrowing. Poplack et al. (2020) provide particularly strong evidence for this distinction by comparing the phonetic treatment of code-switches, nonce borrowings, and established loanwords. They found substantial variability in phonetic adaptation across language-mixing types, in contrast to the immediate and highly consistent morphosyntactic integration observed for borrowed material. Their findings also support the independence of phonological and morphosyntactic integration and show why phonetic realization alone cannot provide a reliable classification of other-language material. In the present corpus, Hungarian-oriented pronunciation and vowel-harmonic suffix selection in all cases where intraword integration occurred strengthen the evidence that English-origin material participates in Hungarian structural processes, but they do not independently establish borrowing status.

In conclusion, morphophonological integration supports the conclusion that the English-origin material participates productively in Hungarian grammatical word formation, but it does not resolve its contact-linguistic classification.

7.9 Code-Switching Versus Borrowing Reconsidered

The corpus contains many forms whose structural composition is transparent but whose status remains theoretically contestable, so the CS vs borrowing problem returns to a central question: what evidence is sufficient to move from grammatical integration to a contact-linguistic classification?

The Poplackian tradition provides an important safeguard against treating lexical origin as a process. Budzhak-Jones (1998), for example, classified English-origin nouns carrying Ukrainian morphology according to their recipient-language behaviour rather than treating the presence of intraword language boundary as code-switching. The Hungarian data are compatible with this logic: an English etymology cannot establish that a form such as *content-et* constitutes a word-internal switch.

At the same time, Hungarian morphology alone cannot establish borrowing within every theoretical framework. MLF explicitly permits English Embedded Language lexical material to receive Hungarian Matrix Language morphology. This is why the same form can support Hungarian structural control without resolving the borrowing – CS distinction.

Deuchar and Stammers (2016) reach a closely related conclusion in their comparison of approaches to English-origin verbs in Welsh. The classification of the same donor-language-origin material changes according to the diagnostics and theoretical definitions employed. Therefore, morphological, syntactic, morphophonological, frequency-based, and dictionary criteria are not interchangeable measures of one underlying category.

The distinction becomes still more difficult when establishment is considered. Dictionary-attested loans were excluded from the present corpus, but dictionary non-attestation cannot establish nonce status, while corpus frequency cannot be equated directly with social establishment. The NBH specifically predicts that grammatical integration can precede widespread recognition, so integration itself cannot be used simultaneously as evidence that a form is both borrowed and established.

Bare items create another problem as they cannot be assumed to be code-switches because they lack visible Hungarian morphology. If the corresponding Hungarian environment itself requires zero marking, the bare English-origin form may be fully compatible with Hungarian grammar, while Owens (2005) also cautions against treating overt morphology as a simple categorical index of integration.

The same caution is needed while analysing multiword material, as Backus (2003) demonstrates that multimorphemic expressions, compounds, collocations, and idioms may function as lexical units. This suggests that the morphologically integrated multiword constructions in the present corpus cannot be classified from word count alone. Some may be complex lexical borrowings, while others may involve productive English internal organization combined with Hungarian external morphology.

However, this does not mean that the borrowing – code-switching distinction should be abandoned. Clear internally English phrases and clauses remain structurally different from lone English-origin bases incorporated into Hungarian morphology. Thus, the findings show that not every naturalistic occurrence contains enough evidence for categorical assignment.

The *indeterminate* category is therefore theoretically legitimate, because it does not refer to an analysis as impossible; rather, that further explanation or expanded specifications are needed from a framework to establish contact status or that further evidence is needed that was not contained in the occurrence itself. Consequently, the most defensible procedure is sequential. Structural properties should be established independently of contact classification first, and then the construction can be evaluated against L^R behaviour, L^D internal structure, evidence of lexicalization, recurrence, and/or community distribution. A borrowing or code-switching label should follow from that evidence rather than determine how the evidence is initially coded.

For the Hungarian-English data, the relevant distinction is between two main questions:

- How is the English-origin material grammatically incorporated?

and

- What evidence justifies classifying that incorporation as code-switching or borrowing?

The corpus answers the first question much more often than it does the second. This reflects the different evidential requirements involved in describing grammar and establishing contact status.

7.10 Implications for Models of Code-Switching

The direct comparison demonstrates that the same bilingual constructions can provide strong evidence for grammatical integration while remaining open to different theoretical classifications. Models of bilingual morphosyntax should therefore distinguish clearly between

structural representation, contact classification, and social establishment, thus avoiding one level to determine the others automatically.

The MLF findings show the value of explicitly representing structural asymmetry. Hungarian grammar control was visible across nominal morphology, verbalization, inflection, category assignment, coverb constructions, and the external morphology of complex English-origin material. At the same time, the model would benefit from increasingly explicit criteria for distinguishing complex EL lexical material from EL islands and for establishing its categories independently of the predictions they are used to test.

On the other hand, the Poplackian analysis demonstrates the importance of L^R benchmarking and of separating grammatical behaviour from superficial foreignness. Its stricter evidential requirements prevent low frequency, English pronunciation, dictionary absence, or multiword status from being treated as automatic classifications. The present data also indicate, however, that morphologically integrated multiword material deserves more explicit treatment, because this category accounted for a substantial portion of the multiword corpus.

A further implication concerns category formation. The Hungarian verbal data show that bilingual integration cannot always be represented as the attachment of inflection to an otherwise unchanged L^D word. English-origin material may first acquire a Hungarian grammatical category through derivation and only then participate in the relevant inflectional paradigm. Models therefore need to distinguish lexical source, category formation, and subsequent inflectional realisation.

The comparison also gives analytical significance to differences in coverage. On the same 405 constructions, the MLF analysis produced a determinate account throughout, whereas 35 constructions remained indeterminate under the operationalized Poplackian analysis. This does not make the MLF Model 8.6 percentage points more “accurate,” but it does show that the model supplied a determinate analysis for a broader portion of this empirical domain. Analytical coverage, diagnostic strength, explicit counterevidence, and indeterminacy should therefore be treated as separate dimensions during framework comparison.

A further implication might be the scalability and reproducibility of Matrix Language identification. Iakovenko and Hain (2024) recently operationalised MLF-informed Matrix Language Identity (MLID) for English-Mandarin and English-Spanish code-switched text and speech and compared these predictions with conventional acoustic language identification.

Their results show that ML-oriented predictors can be evaluated on textual criteria and do not need to rely solely on the acoustic identification of the locally dominant language. This opens a useful direction for extending the present analysis. A computational treatment of Hungarian-English data would require a manually validated training and evaluation corpus together with morphologically informed features capable of representing Hungarian suffixation, verbalization, agreement, case marking, and constituent structure. Such an approach would not replace detailed grammatical analysis, particularly where MOP, 4-M classification, or EL-island status requires linguistic interpretation, but it could test whether the ML coding procedure applied here can be reproduced consistently across other larger datasets.

Finally, models should permit classificatory indeterminacy without structural indeterminacy. A construction may be grammatically well understood even when available evidence does not justify a categorical borrowing or code-switching label. Recognizing this distinction produces a more precise account than forcing every bilingual token into a theoretically predetermined category.

7.11 Methodological Implications of Podcast Data

The use of podcasts has implications beyond the Hungarian-English findings. The corpus shows that publicly available recordings can provide substantial evidence for bilingual morphosyntax under conditions in which the linguistic behaviour was not elicited by the researcher. While podcasts should not be treated as equal to private spontaneous conversation, they are still naturally occurring (media) discourse. The communicative events were produced independently of the research but created for recording and an audience.

7.11.1 Naturalistic Bilingual Interaction

The 24 podcast videos, comprising approximately 30 hours of speech from 37 speakers, existed before and independently of the study. As the speakers were not contacted or recruited to demonstrate code-switching, were not asked to translate expressions, and were not prompted in any way, their bilingual constructions of English material with Hungarian morphology occurred naturally.

This reduces an important form of researcher intervention associated with Labov's (1972) observer's paradox. The speakers knew that they were participating in recorded public interaction, but the specific location and grammatical form of their switches were not responses

to linguistic elicitation. Complex constructions such as *stream-el-ünk*, *feature-e-i-t*, *fake meat-ek-ért*, and *meg akar-ja cheat-el-ni* therefore constitute positive evidence of forms produced without metalinguistic prompting.

Earlier variationist studies relied heavily on conversational corpora, while more recent sources such as the Canberra Vietnamese-English Code-switching Natural Speech Corpus (CanVEC; Nguyen & Bryant, 2020) similarly demonstrate the continuing value of extended natural speech for bilingual corpus research. This value can be found in long-form podcast recordings, which are particularly useful when structurally informative constructions are relatively infrequent. A short elicitation session may produce some intraword switches, but extended recordings allow case marking, possession, verbalization, derivation, mixed compounds, multiword integration, and other configurations to accumulate without targeting them in advance.

7.11.2 Recheckability and Analytical Transparency

Podcast recordings also offer an important practical advantage: the primary audiovisual material can be replayed, and individual constructions can be checked repeatedly in their discourse context, allowing lexical identification, morphological segmentation, constituent boundaries, pronunciation, and glosses to be reconsidered when necessary.

This was especially valuable for ambiguous or morphologically complex forms. The same passage could be (and was) revisited during the theory-neutral coding, MLF analysis, Poplackian analysis, and morphophonological inspection. Classification decisions could therefore be revised without losing access to the original linguistic event.

However, recheckability does not make annotation objective. Decisions about lexical origin, word boundaries, internal constituent structure, or phonological realization still require linguistic judgment. The author acknowledges that manual transcription and glossing carry the potential of making mistakes during the processes. Nevertheless, a stable recording creates a more transparent relationship between the primary data and subsequent analysis and allows questionable decisions to be revisited.

The accessibility of the recordings also improves methodological reproducibility. Where ethically appropriate and legally accessible, other researchers can return to the source material and inspect the context in which the analysed structures occurred without relying solely on decontextualised examples.

7.11.3 Naturalistic Speech as Performed Discourse

Since podcast participants know that they are being recorded and that their interaction will reach an audience, researcher-unelicited speech should not be equated with unmonitored speech. Thus, their linguistic choices may reflect audience design, professional identity, humour, personal branding, topic, or expectations associated with the genre.

This is particularly relevant to their use of English: digital media, entertainment, gaming, business, technology, and other podcast topics represented in the corpus may make English terminology especially accessible or socially appropriate. The frequency and lexical composition of switching in these recordings should not be generalised directly to private or everyday Hungarian conversation.

Podcast formats also differed substantially. Some recordings resembled informal interaction among familiar participants, while others followed presenter – guest structures or discussions on pre-selected topics. *Podcast speech* is therefore not a homogeneous technical term, it simply refers to the source of such corpus.

Editing introduces an additional potential issue with the format, as published recordings may contain cuts or other post-production changes. Such editing is unlikely to generate or modify the morphological constructions analysed here, but it can affect interactional continuity and the observable frequency of hesitations, repairs, repetitions, or abandoned utterances. This means that podcasts are stronger evidence for attested structural possibilities than for precise estimates of their rates in completely unedited speech.

7.11.4 Sampling and Speaker Information

The use of pre-existing recordings also limited speaker metadata. Detailed information concerning age of English acquisition, residence in English-speaking environments, educational history, habitual language use, and language dominance was not available consistently across speakers.

Likewise, English proficiency could not be established through standardised testing. The approximate A2–C1 range used to characterise the participants was inferred from observable performance, and therefore, it should not be used to support claims about relationships between proficiency and particular switching patterns.

The absence of direct contact with the speakers also prevented follow-up questions about individual expressions. Speakers could not be asked whether an English-origin form was familiar, newly created, frequently used, deliberately humorous, or perceived as Hungarian or English. Information of this nature would be especially useful for the borrowing - code-switching distinction because lexical familiarity and community establishment cannot be reconstructed reliably from surface form alone.

The sample is therefore not representative of the Hungarian-speaking population. Podcast speakers form a self-selecting group participating in public media interaction and may differ systematically from other speakers.

7.11.5 Implications for Corpus Interpretation

The evidential value of podcast data depends on the question asked. They are strong for identifying positive structural evidence, as if an English-origin base is attested with Hungarian case, verbalization, agreement, or coverb syntax, the recording demonstrates that this configuration was produced in naturalistic interaction. Although, non-occurrence is much weaker evidence. The absence of a particular construction cannot establish ungrammaticality because the relevant lexical item, syntactic environment, or communicative context may simply not have arisen. Experimental elicitation or acceptability judgments remain necessary where the research question concerns grammatical impossibility or the responses of the speakers to tightly controlled contrasts.

Podcast and experimental data should therefore be viewed as complementary. Experimental methods provide control over proficiency, lexical frequency, structural variables, and participant background. Naturalistic podcast data show which constructions speakers actually do produce when the researcher does not determine when or how switching should occur.

Public accessibility also raises ethical considerations. As mentioned earlier, publicly available speech should not automatically be treated as unrestricted research material simply because it can be accessed online. BAAL (2021) emphasizes contextual integrity, identifiability, and potential harm, while Stommel and de Rijk (2021) show that research using publicly available online discourse still requires explicit consideration of how participants, consent, anonymity, and retrievability are handled. In the present study, speaker identities were not analytically necessary and were therefore not included in the dissertation, while also being anonymised in the publicly available data.

Therefore, podcast corpora occupy a useful methodological position between researcher-created experimental data and inaccessible private interaction. Their strengths are researcher-independent production, contextual richness, duration, and recheckability and their limitations concern representativeness, performance, editing, speaker metadata, and the absence of direct evidence about lexical knowledge or community establishment. For this dissertation, those strengths were particularly well matched to the research questions. The recordings provided direct evidence of how Hungarian-English morphosyntactic integration occurs in extended public interaction without being requested by the researcher.

7.12 Limitations

Among the several limitations that qualify the interpretation of the findings, probably the nature of the data provides the most. First, the corpus was constructed from publicly available recordings. This provided researcher-unelicited bilingual speech but limited the availability of demographic and language-background information: proficiency, language dominance, age of acquisition, and habitual English use could not be measured systematically.

Second, podcast speech is naturalistic but publicly performed. It should be noted that the topic, audience, professional identity, genre expectations, and editing may influence both lexical choice and switching behaviour. The corpus therefore cannot be treated as representative of everyday Hungarian conversation, even though, many observable discussions contained explicit language (some examples of which are present in the data) or of the Hungarian-speaking population more generally.

Also because of the nature of the data, the corpus provides only positive grammatical evidence. Attested constructions establish that certain combinations occur, but non-attestation does not establish ungrammaticality.

Next, the borrowing – code-switching distinction could not be resolved for every construction. Dictionary non-attestation and low corpus frequency were insufficient to establish nonce status, while information concerning familiarity and community distribution was unavailable. This limitation was most visible in the 35 morphologically integrated multiword constructions, which remained indeterminate under the Poplackian analysis. For the same framework, the study demonstrated extensive Hungarian morphosyntactic integration, however it did not conduct a full variationist comparison with matched native Hungarian forms across equivalent variable environments. Still, the data are compatible with the NBH, including the

possibility of immediate L^R integration, but cannot establish complete distributional equivalence between all English-origin and native Hungarian items.

Importantly, the 100% MLF and 91.4% Poplackian coverage figures are meaningfully comparable as measures of analytical coverage across the same 405 constructions, but they should not be interpreted as statistical accuracy scores. Some constructions directly tested language-specific predictions, whereas others received determinate analyses despite being less diagnostic. Additionally, MLF categories such as EL islands and complex ML+EL constituents must be established independently to avoid circularity.

Finally, the morphophonological findings were based primarily on repeated auditory examination rather than systematic acoustic analysis, and the earlier detailed phonological classification was conducted on a smaller version of the intraword corpus (although not by much). These results therefore support a strong tendency toward Hungarian-oriented phonological accommodation but cannot establish the precise phonological representation of every mixed form.

These limitations limit claims about population frequency, speaker cognition, lexical establishment, categorical grammatical impossibility, and the relative statistical performance of the frameworks, but they do not alter the main findings of the corpus.

8. CONCLUSIONS

This dissertation examined the morphosyntactic integration of English-origin material in naturally occurring Hungarian podcast speech and compared the same constructions under the Matrix Language Frame (MLF) Model and the Poplackian variationist approach, including the Nonce Borrowing Hypothesis (NBH). The corpus contained 419 bilingual forms, of which 405 intraclausal constructions were retained for structural analysis.

The findings show a highly consistent pattern of Hungarian grammatical organization. English supplied primarily lexical material, whereas Hungarian provided category formation, inflection, case, possession, verbal morphology, and external constituent structure. At the same time, the analysis demonstrated that grammatical integration can often be established more confidently than the resulting form can be classified as code-switching or borrowing.

8.1 Answers to the Research Questions

The first research question asked how English-origin material is integrated into Hungarian morphosyntax. As presented through the data, integration occurred across nominal, verbal, and adjectival domains and included case, number, possession, verbalization, adjectival derivation, comparison, nominalization, verbal inflection, and coverb constructions. Of the 405 intraclausal constructions, 187 contained intraword integration, including 152 lone items and 35 morphologically integrated multiword constructions. All 40 English-origin verbal constructions contained Hungarian verbalizing morphology, confirming and extending the integration pattern described by Bolonyai (2005).

The second research question concerned the MLF Model. In its analysis, Hungarian was identifiable as the Matrix Language throughout the intraclausal dataset, and no unambiguous violation of the applicable Morpheme Order or System Morpheme predictions was identified. Diagnostic intraword constructions followed Hungarian morpheme order, the relevant outsider late system morphemes originated in Hungarian, and internally English sequences could be accommodated as EL islands where English internal grammar was independently identifiable. The MLF Model provided a determinate structural analysis for all 405 constructions, although not every construction was equally diagnostic. Regardless of lacking equal diagnostics in all examples, the first prediction concerning the MLF is supported, as no violations of the model occurred.

The third research question concerned the Poplackian approach and the NBH. The 152 overtly integrated lone English-origin items displayed clear Hungarian morphosyntax, while the clearest multiword English sequences retained English internal structure. Therefore, the data were compatible with the Poplackian distinction between recipient-language-integrated borrowing and donor-language-structured switching (Poplack, 2018). They were also compatible with the NBH prediction that grammatical integration does not require prior social establishment (Sankoff et al., 1990; Poplack, 2012; Poplack & Dion, 2012). However, nonce status could not be demonstrated independently for individual items, and no complete matched variationist comparison with native Hungarian forms was conducted. Also, the Poplackian analysis provided a determinate account for 370 constructions, while 35 morphologically integrated multiword constructions remained indeterminate. Accordingly, the second prediction concerning the classification of the Poplackian approach is only partially supported, as multiword code-switches still received Hungarian morphology.

The fourth research question asked where the frameworks converge and diverge. Both approaches agreed that Hungarian supplied the relevant external grammatical structure and that clearly structured English multiword sequences could preserve English internal grammar. Their main disagreement concerned the interpretation of integrated English-origin material. Under the MLF, such material may remain EL material within a Hungarian ML frame; under the Poplackian approach, L^R-equivalent grammatical behaviour is borrowing-compatible. The difference between 100% MLF coverage and 91.4% Poplackian coverage should be interpreted as a difference in their success in analytical coverage. This finding supports the third prediction concerning constituent size introduced in Section 1.6

Finally, the last prediction on morphophonological integration (which is expected to serve as evidence of Hungarian grammatical control but not in determining borrowing status independently) is also supported, as Hungarian vowel harmony, suffix selection, and phonological accommodation accompanied morphosyntactic integration without providing a categorical boundary between code-switching and borrowing.

8.2 Main Theoretical Conclusions

The main theoretical conclusion is that morphosyntactic integration does not independently determine contact status. Hungarian morphology can establish how English-origin material

participates in the grammatical structure without deciding whether the resulting construction should be labelled as code-switching or borrowing.

A second conclusion is that constituent size and grammatical integration are independent dimensions. The 35 morphologically integrated multiword constructions show that complex English-origin material can receive Hungarian morphology externally. So, multiword status alone does not establish productive English syntax or structural independence from Hungarian. This is consistent with evidence that multimorphemic and multiword material may function as lexical units in bilingual speech (Backus, 2003).

The data also show that bilingual integration may involve category formation before inflection. English-origin material can first acquire a Hungarian verbal, adjectival, or nominal category and then enter the relevant inflectional grammar. This is especially clear in the verbal constructions, where Hungarian verbalizers precede subsequent grammatical marking.

Finally, the comparison supports a distinction between structural indeterminacy and classificatory indeterminacy. A construction may be grammatically transparent even where its borrowing/code-switching status cannot be established. Similarly, analytical coverage and diagnostic strength should be treated separately: a construction may receive a determinate analysis without constituting an equally strong test of every theoretical prediction.

Morphophonological evidence supports these conclusions without resolving the classification problem. Hungarian-oriented pronunciation, vowel harmony, and suffix selection reinforce the evidence for Hungarian structural participation, but phonological and morphosyntactic integration are not equivalent diagnostics of borrowing status (Poplack et al., 2020).

8.3 Contributions of the Dissertation

8.3.1 Empirical Contribution

The study contributes a detailed account of Hungarian-English morphosyntactic integration based on approximately 30 hours of naturalistic podcast speech involving 37 speakers. It documents integration across nominal, verbal, adjectival, possessive, derivational, and syntactic domains not only concentrating on a single grammatical category.

The 35 morphologically integrated multiword constructions are particularly significant because they show that Hungarian morphology can operate on complex English-origin bases, so their presence demonstrates that integration is not restricted to lone lexical items.

8.3.2 Theoretical Contribution

The main theoretical contribution lies in applying independently operationalised predictions of the MLF and Poplackian approaches to the same dataset. Doing so makes it possible to distinguish agreement about observable structure from disagreement about theoretical classification.

The study also shows that analytical coverage is a meaningful dimension of framework comparison. MLF provided a determinate structural analysis throughout the corpus, whereas the Poplackian approach left 35 constructions indeterminate. This difference does not necessarily establish general theoretical superiority, but it identifies where the frameworks differ in the extent to which they classify the same empirical domain.

More broadly, the dissertation demonstrates the value of Hungarian as a test case for bilingual morphosyntax. Its overt case marking, derivation, verbal agreement, possession, and coverb system make grammatical integration clearly visible and allow individual theoretical predictions to be tested directly.

8.3.3 Methodological Contribution

Methodologically, the study supports the claim about the usefulness of podcasts as sources of researcher-unelicited bilingual speech. Online available video recordings provide extended interaction, preserve the original discourse context, and allow constructions to be repeatedly checked during transcription and analysis.

Another methodological contribution is the use of theory-neutral descriptive coding before framework-specific classification. English-origin material was first described according to its observable structural properties, reducing the risk that the assumptions of either framework determined the initial classification of the data.

8.4 Directions for Future Research

First of all, future research should strengthen the Poplackian comparison by examining English-origin and native Hungarian items in matched grammatical environments. Community-based

data and more detailed speaker information would also allow more reliable investigation of lexical establishment, proficiency, dominance, and familiarity with individual English expressions.

The 35 morphologically integrated multiword constructions deserve targeted study through recurrence data, prosodic analysis, speaker judgments, and experimental elicitation. Such work could distinguish more reliably between compounds, stored multiword units, and productively generated phrases.

Acoustic analysis could further test morphophonological evidence by comparing mixed forms with the speakers' Hungarian and English productions. This would refine the description of vowel quality, phonological accommodation, and suffix harmony without treating phonological integration as an independent borrowing criterion.

A further direction concerns the scalability of ML identification. Building on Iakovenko and Hain's (2024) work on automatic MLID determination, future Hungarian-English research could test whether manually established classifications can be reproduced computationally on a larger corpus using morphologically informed features representing Hungarian suffixation, verbalization, agreement, case marking, and constituent structure.

Finally, this analytical procedure should be applied to other morphologically rich language pairs, which would help establish which patterns reflect broader properties of bilingual morphosyntactic integration and which are specific to Hungarian-English contact.

References

Akadémiai Kiadó. (n.d.). *Magyar értelmező kéziszótár* [*Concise Explanatory Dictionary of Hungarian*]. szotar.net. Retrieved Feb 3, 2026.

Alexiadou, A. (2020). Compound formation in language mixing. *Frontiers in Psychology, 11*, Article 1021. https://doi.org/10.3389/fpsyg.2020.01021

Alexiadou, A. (2024). Double plural marking and the building blocks of nominals. *Glossa: A Journal of General Linguistics, 9*(1), 1–35. https://doi.org/10.16995/glossa.11585

Auer, P. (Ed.). (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge.

Backus, A. (1992). *Patterns of language mixing: A study in Turkish–Dutch bilingualism*. Harrassowitz.

Backus, A. (1996). *Two in one: Bilingual speech of Turkish immigrants in the Netherlands*. Tilburg University Press.

Backus, A. (2003). Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics, 41*(1), 83–132. https://doi.org/10.1515/ling.2003.005

Backus, A. (2009). Codeswitching as one piece of the puzzle of language change: The case of Turkish *yapmak*. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary approaches to code switching* (pp. 307–336). John Benjamins. https://doi.org/10.1075/sibil.41.16bac

Backus, A. (2014). A usage-based approach to borrowability. In E. Zenner & G. Kristiansen (Eds.), *New perspectives on lexical borrowing: Onomasiological, methodological and phraseological innovations* (pp. 19–40). De Gruyter Mouton. https://doi.org/10.1515/9781614514305.19

Backus, A., & Demirçay, D. (2021). Intense Turkish–Dutch bilingualism leads to intense Turkish–Dutch mixing: A usage-based account of increasing integration of two typologically different languages. *Belgian Journal of Linguistics, 35*, 13–33. https://doi.org/10.1075/bjl.00062.bac

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(6), 1022–1047. https://doi.org/10.1037/xlm0000770

Belazi, H. M., Rubin, E. J., & Toribio, A. J. (1994). Code switching and X-bar theory: The Functional Head Constraint. *Linguistic Inquiry, 25*(2), 221–237.

Berk-Seligson, S. (1986). Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society, 15*(3), 313–348. https://doi.org/10.1017/S0047404500011799

Bhatt, R. M., & Bolonyai, A. (2011). Code-switching and the optimal grammar of bilingual language use. *Bilingualism: Language and Cognition, 14*(4), 522–546. https://doi.org/10.1017/S1366728910000295

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *The Journal of Neuroscience, 37*(37), 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017

Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*(12), 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001

Bolonyai, Á. (2005). English verbs in Hungarian/English codeswitching. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 317–327). Cascadilla Press.

British Association for Applied Linguistics. (2021). *Recommendations on good practice in applied linguistics* (4th ed.). https://www.baal.org.uk/wp-content/uploads/2021/03/BAAL-Good-Practice-Guidelines-2021.pdf

Budzhak-Jones, S. (1998). Against word-internal codeswitching: Evidence from Ukrainian–English bilingualism. *International Journal of Bilingualism, 2*(2), 161–182. https://doi.org/10.1177/136700699800200203

Bullock, B. E., & Toribio, A. J. (Eds.). (2009). *The Cambridge handbook of linguistic code-switching*. Cambridge University Press.

Chi, J., & Bell, P. (2024). Analyzing the role of part-of-speech in code-switching: A corpus-based study. In Y. Graham & M. Purver (Eds.), *Findings of the Association for Computational Linguistics: EACL 2024* (pp. 1712–1721). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-eacl.120

Comrie, B., Haspelmath, M., & Bickel, B. (2008). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. Max Planck Institute for Evolutionary Anthropology. https://www.eva.mpg.de/lingua/resources/glossing-rules.php

Coppock, E. (2013). A semantic solution to the problem of Hungarian object agreement. *Natural Language Semantics, 21*, 345–371. https://doi.org/10.1007/s11050-013-9096-7

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

de Groot, C. (2005). The grammars of Hungarian outside Hungary from a linguistic-typological perspective. In A. Fenyvesi (Ed.), *Hungarian language contact outside Hungary: Studies on Hungarian as a minority language* (pp. 351–370). John Benjamins. https://doi.org/10.1075/impact.20.18gro

Demirçay, D., & Backus, A. (2014). Bilingual constructions: Reassessing the typology of code-switching. *Dutch Journal of Applied Linguistics, 3*(1), 30–44. https://doi.org/10.1075/dujal.3.1.02dem

Demszky, D. (2021). The role of verb semantics in Hungarian verb-object order. *Proceedings of the Linguistic Society of America, 6*(1), 54–68. https://doi.org/10.3765/plsa.v6i1.4941

Deuchar, M. (2005). Congruence and Welsh–English code-switching. *Bilingualism: Language and Cognition, 8*(3), 255–269. https://doi.org/10.1017/S1366728905002294

Deuchar, M. (2006). Welsh–English code-switching and the Matrix Language Frame model. *Lingua, 116*(11), 1986–2011. https://doi.org/10.1016/j.lingua.2005.06.001

Deuchar, M. (2012). Code switching. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 657–664). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0142

Deuchar, M. (2020). Code-switching in linguistics: A position paper. *Languages, 5(2),* Article 22. https://doi.org/10.3390/languages5020022

Deuchar, M., Davies, P., Herring, J. R., Parafita Couto, M. C., & Carter, D. (2014). Building bilingual corpora. In E. M. Thomas & I. Mennen (Eds.), *Advances in the study of bilingualism* (pp. 93–110). Multilingual Matters. https://doi.org/10.21832/9781783091713-008

Deuchar, M., & Stammers, J. R. (2012). What is the "Nonce Borrowing Hypothesis" anyway? *Bilingualism: Language and Cognition, 15*(3), 649–650. https://doi.org/10.1017/S1366728911000563

Deuchar, M., & Stammers, J. R. (2016). English-origin verbs in Welsh: Adjudicating between two theoretical approaches. *Languages, 1*(1), Article 7. https://doi.org/10.3390/languages1010007

Di Sciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics, 22*(1), 1–24. https://doi.org/10.1017/S0022226700010537

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Doğruöz, A. S., Sitaram, S., Bullock, B. E., & Toribio, A. J. (2021). A survey of code-switching: Linguistic and social perspectives for language technologies. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.),

Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1654–1666). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.131

Domokos, J., & Deganutti, M. (2021). Four major literary code-switching strategies in Hungarian literature: Decoding monolingualism. *Hungarian Studies Yearbook, 3*(1), 43–63. https://doi.org/10.2478/hsy-2021-0004

Duran Eppler, E., Luescher, A., & Deuchar, M. (2017). Evaluating the predictions of three syntactic frameworks for mixed determiner–noun constructions. *Corpus Linguistics and Linguistic Theory, 13*(1), 27–63. https://doi.org/10.1515/cllt-2015-0006

É. Kiss, K. (1987). *Configurationality in Hungarian*. Akadémiai Kiadó.

É. Kiss, K. (Ed.). (1995). *Discourse configurational languages*. Oxford University Press.

É. Kiss, K. (1998). Identificational focus versus information focus. *Language, 74*(2), 245–273. https://doi.org/10.2307/417867

É. Kiss, K. (2002). *The syntax of Hungarian*. Cambridge University Press.

Farkas, J., & Kniezsa, V. (2004). Hungarian. In M. Görlach (Ed.), *English in Europe* (pp. 277–290). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199273102.003.0015

Fenyvesi, A. (2005a). Hungarian in the United States. In A. Fenyvesi (Ed.), *Hungarian language contact outside Hungary: Studies on Hungarian as a minority language* (pp. 265–318). John Benjamins. https://doi.org/10.1075/impact.20.16fen

Fenyvesi, A. (Ed.). (2005b). *Hungarian language contact outside Hungary: Studies on Hungarian as a minority language*. John Benjamins. https://doi.org/10.1075/impact.20

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge University Press.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Blackwell.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 21–39). Cambridge University Press.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.

Halmari, H. (1997). *Government and codeswitching: Explaining American Finnish*. John Benjamins. https://doi.org/10.1075/sibil.12

Han, X., Li, W., & Filippi, R. (2022). The effects of habitual code-switching in bilingual language production on cognitive control. *Bilingualism: Language and Cognition, 25(5),* 869–889. https://doi.org/10.1017/S1366728922000244

Haugen, E. (1950). The analysis of linguistic borrowing. *Language, 26*(2), 210–231. https://doi.org/10.2307/410058

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. C. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language, 85*(4), 822–863. https://doi.org/10.1353/lan.0.0169

Hlavac, J. (1999). Phonological integration of English transfers in Croatian: Evidence from the Croatian speech of second-generation Croatian-Australians. *Filologija, 32*, 39–74.

Hlavac, J. (2003). *Second-generation speech: Lexicon, code-switching and morpho-syntax of Croatian–English bilinguals*. Peter Lang.

Iakovenko, O., & Hain, T. (2024). Methods of automatic Matrix Language determination for code-switched speech. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 5791–5800). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.330

Jake, J. L., & Myers-Scotton, C. (1997). Codeswitching and compromise strategies: Implications for lexical structure. *International Journal of Bilingualism, 1*(1), 25–39. https://doi.org/10.1177/136700699700100103

Jake, J. L., Myers-Scotton, C., & Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the Matrix Language. *Bilingualism: Language and Cognition, 5*(1), 69–91. https://doi.org/10.1017/S1366728902000147

Jake, J. L., Myers-Scotton, C., & Gross, S. (2005). A response to MacSwan (2005): Keeping the Matrix Language. *Bilingualism: Language and Cognition, 8*(3), 271–276. https://doi.org/10.1017/S1366728905002300

Johns, M. A., & Dussias, P. E. (2022). Comparing single-word insertions and multi-word alternations in bilingual speech: Insights from pupillometry. *Languages, 7*(4), Article 267. https://doi.org/10.3390/languages7040267

Johns, M. A., & Steuck, J. (2021). Is codeswitching easy or difficult? Testing processing cost through the prosodic structure of bilingual speech. *Cognition, 211*, Article 104634. https://doi.org/10.1016/j.cognition.2021.104634

Kalivoda, Á. (2022). PrevDistro: An open-access dataset of Hungarian preverb constructions. *Acta Linguistica Academica, 69*(4), 549–563. https://doi.org/10.1556/2062.2022.00578

Kardos, É., & Farkas, I.-Á. (2022). The syntax of inner aspect in Hungarian. *Journal of Linguistics, 58*(4), 807–845. https://doi.org/10.1017/S0022226721000426

Kenesei, I., Vago, R. M., & Fenyvesi, A. (1998). *Hungarian*. Routledge.

Kiefer, F. (Ed.). (2000). *Strukturális magyar nyelvtan 3: Morfológia* [Structural grammar of Hungarian 3: Morphology]. Akadémiai Kiadó.

Kiefer, F. (2012). Uralic, Finno-Ugric: Hungarian. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp. 527–541). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199695720.013.0029

Kontra, M. (Ed.). (1995). *Túl a Kecegárdán/Beyond Castle Garden: An American dictionary of the Calumet region* (A. Vázsonyi, Comp.). Teleki László Alapítvány. Budapest

Kovács, M. (2005). Hungarian in Australia. In A. Fenyvesi (Ed.), *Hungarian language contact outside Hungary: Studies on Hungarian as a minority language* (pp. 319–350). John Benjamins. https://doi.org/10.1075/impact.20.17kov

Kovács, T. (2011). How Optimality Theory works for bilingual grammar: On the applicability of Optimality Theory to bilingual grammar in the description of English-Hungarian code-switching patterns. *AMERICANA E-Journal of American Studies in Hungary, 7*. https://americanaejournal.hu/index.php/americanaejournal/article/view/45246

Kovács, T. (2018). *Code-switching and optimality: An Optimality-Theoretical approach to the socio-pragmatic patterns of Hungarian-English code-switching*. L'Harmattan. https://doi.org/10.56037/978-2-343-15050-5

Kovács, T. (2024). The meaning of code-switching in Anne Tardos's multilingual poem *Minden Ami*. *Orpheus Noster, 16*(3), 12–23.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

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

Lewis, G., Jones, B., & Baker, C. (2012). Translanguaging: Origins and development from school to street and beyond. *Educational Research and Evaluation, 18*(7), 641–654. https://doi.org/10.1080/13803611.2012.718488

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30. https://doi.org/10.1093/applin/amx039

MacSwan, J. (1999). *A minimalist approach to intrasentential code switching*. Garland.

MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition, 3*(1), 37–54. https://doi.org/10.1017/S1366728900000122

MacSwan, J. (2005). Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism." *Bilingualism: Language and Cognition, 8*(1), 1–22. https://doi.org/10.1017/S1366728904002068

MacSwan, J. (Ed.). (2014). *Grammatical theory and bilingual codeswitching*. MIT Press.

MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal, 54*(1), 167–201. https://doi.org/10.3102/0002831216683935

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

Mahootian, S. (1993). *A null theory of code-switching* [Doctoral dissertation, Northwestern University].

Mahootian, S., & Santorini, B. (1996). Code switching and the complement/adjunct distinction. *Linguistic Inquiry, 27*(3), 464–479.

Meechan, M., & Poplack, S. (1995). Orphan categories in bilingual discourse: Adjectivization strategies in Wolof–French and Fongbe–French. *Language Variation and Change, 7*(2), 169–194. https://doi.org/10.1017/S0954394500000971

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Milroy, L., & Muysken, P. (Eds.). (1995). *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge University Press.

Moravcsik, E. A. (1975). Verb borrowing. *Wiener Linguistische Gazette, 8*, 3–30.

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

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

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299530.001.0001

Myers-Scotton, C., & Bolonyai, A. (2001). Calculating speakers: Codeswitching in a rational choice model. *Language in Society, 30*(1), 1–28. https://doi.org/10.1017/S0047404501001014

Myers-Scotton, C., & Jake, J. L. (2000). Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics, 38*(6), 1053–1100. https://doi.org/10.1515/ling.2000.021

Myers-Scotton, C., & Jake, J. L. (2001). Explaining aspects of code-switching and their implications. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 84–116). Blackwell.

Myers-Scotton, C., & Jake, J. L. (2014). Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition, 17*(3), 511–525. https://doi.org/10.1017/S1366728913000758

Navracsics, J. (2007a). *A kétnyelvű mentális lexikon* [The bilingual mental lexicon]. Balassi Kiadó.

Navracsics, J. (2007b). Word classes and the bilingual mental lexicon. In Zs. Lengyel & J. Navracsics (Eds.), *Second language lexical processes: Applied linguistic and psycholinguistic perspectives* (pp. 17–36). Multilingual Matters. https://doi.org/10.21832/9781853599682-004

Navracsics, J. (2010). Kódváltás és kódkeverés kétnyelvűek spontán beszédében [Code-switching and code-mixing in bilinguals' spontaneous speech]. In J. Navracsics (Ed.), *Nyelv, beszéd, írás: Pszicholingvisztikai tanulmányok I* [Language, speech, writing: Psycholinguistic studies I] (pp. 121–129). Tinta Könyvkiadó.

Navracsics, J. (2011). *Szóaktiváció két nyelven* [Word activation in two languages]. Gondolat Kiadó.

Nguyen, L. (2025). Rethinking the matrix language: Vietnamese–English code-switching in Canberra. *International Journal of Bilingualism, 29*(4), 1067–1093. https://doi.org/10.1177/13670069241254454

Nguyen, L., & Bryant, C. (2020). CanVEC – The Canberra Vietnamese-English code-switching natural speech corpus. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 4121–4129). European Language Resources Association.

Otheguy, R., García, O., & Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review, 6*(3), 281–307. https://doi.org/10.1515/applirev-2015-0014

Owens, J. (2005). Bare forms and lexical insertions in code-switching: A processing-based account. *Bilingualism: Language and Cognition, 8*(1), 23–38. https://doi.org/10.1017/S1366728904002056

Özkara, B., Cedden, G., von Stutterheim, C., & Meyer, P. (2025). Code-switching and cognitive control: A review of current trends and future directions. *Frontiers in Language Sciences, 4,* Article 1515283. https://doi.org/10.3389/flang.2025.1515283

Parafita Couto, M. C., & Gullberg, M. (2019). Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism, 23*(2), 695–714. https://doi.org/10.1177/1367006917729543

Pierantozzi, C., & Cocchi, G. (2024). Mixed compounds in code-switching contexts. *LingBaW: Linguistics Beyond and Within, 10*, 177–193. https://doi.org/10.31743/lingbaw.18018

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. https://doi.org/10.1515/ling.1980.18.7-8.581

Poplack, S. (2012). What does the Nonce Borrowing Hypothesis hypothesize? *Bilingualism: Language and Cognition, 15*(3), 644–648. https://doi.org/10.1017/S1366728911000496

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford University Press.

Poplack, S., & Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change, 24*(3), 279–315. https://doi.org/10.1017/S095439451200018X

Poplack, S., & Meechan, M. (1998). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism, 2*(2), 127–138. https://doi.org/10.1177/136700699800200201

Poplack, S., Robillard, S., Dion, N., & Paolillo, J. C. (2020). Revisiting phonetic integration in bilingual borrowing. *Language, 96*(1), 126–159. https://doi.org/10.1353/lan.2020.0004

Poplack, S., & Sankoff, D. (1984). Borrowing: The synchrony of integration. *Linguistics, 22*(1), 99–135. https://doi.org/10.1515/ling.1984.22.1.99

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics, 26*(1), 47–104. https://doi.org/10.1515/ling.1988.26.1.47

Poplack, S., Wheeler, S., & Westwood, A. (1989). Distinguishing language contact phenomena: Evidence from Finnish–English bilingualism. *World Englishes, 8*(3), 389–406. https://doi.org/10.1111/j.1467-971X.1989.tb00677.x

Rebrus, P., & Törkenczy, M. (2021). Harmonic Uniformity and Hungarian front/back harmony. *Acta Linguistica Academica, 68*(1–2), 175–206. https://doi.org/10.1556/2062.2021.00475

Salamon, A., & Szentgyörgyi, Sz. (2026). *Hungarian-English intraword code-switching* [Manuscript submitted for publication].

Sankoff, D., & Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics, 14*(1), 3–45. https://doi.org/10.1080/08351818109370523

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change, 2*(1), 71–101. https://doi.org/10.1017/S0954394500000272

Shaw, M., & De Smet, H. (2022). Loan word accommodation biases: Markedness and finiteness. *Transactions of the Philological Society, 120*(2), 201–217. https://doi.org/10.1111/1467-968X.12233

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. Oxford University Press.

Stammers, J. R., & Deuchar, M. (2012). Testing the Nonce Borrowing Hypothesis: Counter-evidence from English-origin verbs in Welsh. *Bilingualism: Language and Cognition, 15*(3), 630–643. https://doi.org/10.1017/S1366728911000381

Stefanich, S., Cabrelli, J., Hilderman, D., & Archibald, J. (2019). The morphophonology of intraword codeswitching: Representation and processing. *Frontiers in Communication, 4*, Article 54. https://doi.org/10.3389/fcomm.2019.00054

Stommel, W., & de Rijk, L. (2021). Ethical approval: None sought. How discourse analysts report ethical issues around publicly available online data. *Research Ethics, 17*(3), 275–297. https://doi.org/10.1177/1747016120988767

Surányi, B. (2009). Verbal particles inside and outside vP. *Acta Linguistica Hungarica, 56*(2–3), 201–249. https://doi.org/10.1556/ALing.56.2009.2-3.3

Treffers-Daller, J. (2005). Evidence for insertional codemixing: Mixed compounds and French nominal groups in Brussels Dutch. *International Journal of Bilingualism, 9*(3–4), 477–506. https://doi.org/10.1177/13670069050090030901

Treffers-Daller, J. (2025). The Simple View of borrowing and code-switching. *International Journal of Bilingualism, 29*(2), 347–370. https://doi.org/10.1177/13670069231168535

van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62*, 148–171. https://doi.org/10.1111/j.1467-9922.2012.00710.x

Vishogradska, I. (2017). Marking the foreign word – A case study of recent English loans in Hungarian. *Acta Universitatis Sapientiae, Philologica, 9*(2), 53–60. https://doi.org/10.1515/ausp-2017-0017

Weinreich, U. (1953). *Languages in contact: Findings and problems*. Linguistic Circle of New York.

Wilson, D. V., & Dumont, J. (2015). The emergent grammar of bilinguals: The Spanish verb *hacer* 'do' with a bare English infinitive. *International Journal of Bilingualism, 19*(4), 444–458. https://doi.org/10.1177/1367006913516047

Wohlgemuth, J. (2009). *A typology of verbal borrowings*. De Gruyter Mouton. https://doi.org/10.1515/9783110219340

Zabrodskaja, A. (2009). Evaluating the Matrix Language Frame model on the basis of a Russian-Estonian codeswitching corpus. *International Journal of Bilingualism, 13*(3), 357–377. https://doi.org/10.1177/1367006908346629

Zabrodskaja, A., & Verschik, A. (2014). Morphology of Estonian items at the interface of Russian-Estonian language contact data. *Sociolinguistic Studies, 8*(3), 449–474.

Appendices

Appendix A - Glossing Abbreviations

N-noun

V-verb

ADJ-adjective

ADV-adverb

GER-gerund

NUM-numeral

PRON-pronoun

CONJ- conjunction

DET-determiner

AUX-auxiliary verb

MOD-modal verb

COV-coverb

PERF-perfective

POSTPOS-postposition

PREP-preposition

NEG-negative

1S-first person, singular

2S-second person, singular

3S-third person, singular

1PL-first person, plural

2PL-second person, plural

3PL-third person, plural

PAST-past tense

PR-present tense

FUT-future tense

NOM-nominal

ACC-accusative

DEM-demonstrative

IMP-imperative

COMP-comparative

SUP-superlative

DAT-dative

INF-infinitive

INTER-interrogative

INTRANS-intransitive

COND-conditional

POT-potential

INSTR-instrumental

REL-relative

POSS-possessive

TERM-terminative

CAUS-FIN-cause-final

DEF-definite

INDEF-indefinite

NZ-nominalizer

ADJZ-adjectivizer

VBZ-verbalizer

ADVZ-adverbializer

IM-interaction marker

DM-discourse marker

UNI-universal (pronoun)

SUBLAT-sublative

ELAT-elative

SUPERESS- superessive

ILLAT-illative

ABLAT-ablative

INESS-inessive

DELAT-delative

ALL-allative

ADESS-adessive

FREQ-frequentative

CAUS-causative

ESS-essive

TRANSLAT-translative

IJ-interjection

REFL-reflexive

Appendix B - Podcast Sources

https://www.youtube.com/watch?v=1KXVCiuE4IQ

https://www.youtube.com/watch?v=4c4cRxG1DS4

https://www.youtube.com/watch?v=dJppPnb8EPc

https://www.youtube.com/watch?v=RRvhEuwzOno

https://www.youtube.com/watch?v=8lZ7HLRk1Rg

https://www.youtube.com/watch?v=VZnn04aTTz0

https://www.youtube.com/watch?v=0K3znIYDze0

https://www.youtube.com/watch?v=OT5XyEyfOHA

https://youtu.be/UtxYRWbx3MM

https://youtu.be/11BiQyEcdXw

https://www.youtube.com/watch?v=LxuNllV7NjA

https://www.youtube.com/watch?v=i0czvn2-aiM

https://www.youtube.com/watch?v=ctMBjyLmeXQ

https://www.youtube.com/watch?v=gPZcFmX057c

https://youtu.be/FiXHgJ_rHXM

https://www.youtube.com/watch?v=TOn7__Y40xM

https://www.youtube.com/watch?v=M74ENNqWASU

https://www.youtube.com/watch?v=aCtyZ8xw8To

https://www.youtube.com/watch?v=ocl8ZxT4jVo

https://youtu.be/hAykeKjU-H0

https://www.youtube.com/watch?v=KVVIHZHlkcM

www.youtube.com/watch?v=M2G6whnAApI

https://www.youtube.com/watch?v=t_jksGcEZSQ

https://www.youtube.com/watch?v=0iXsPY9qPkw

Appendix C - Publicly Available Datasheet of the Corpus

https://osf.io/kzfah/overview?view_only=178b38783cea4428be352309e5e7bc81

Appendix D - Morphologically Integrated Multiword Code-Switching Tokens

Intraword + Multiword
de én csak ilyen sleep tracking-re használ-om
pontos sleep tracking-et csak agy-hullám-ból tud-sz (kapni)
és utána megnéz-t-em, hogy milyen sleep tracker-ek van-nak a piac-on
száz-ezer concurrent viewer-rel akar-sz ki-vonul-ni a világ-ból
van olyan a YouTube-on aki így side project-be csinál olyat ami saját platform
en-ni kéne a tide pod-ot
én is mindig beszél-ek a trash content-ről
egy live stream-be' ugye én egy-be rak-t-am a 10 szám-ot
amikor a live stream-ek-et el-kezd-t-ük
a leg-durvá-bb community guideline set-tel rendelkez-ő platform a TikTok
lehet kér-ni join request-et
amikor volt gaming channel-öm
egész délután-ok-at gyakorol-t-am a heel flip-et
a frech fries-nak is valami ilyesmi a gyöker-e
a french fries-t nem volt szabad mond-ani
ez most valamilyen szint-en a cancel culture-t fel-vált-ott-a
ő private jet-tel jár Párizs-ba reggeli-z-ni
a feel good-ok az-ok még hiányoz-nak
régen jár-t-am ilyen jam session-ök-be
mindenki úgy néz-ett ki mintha valami fashion week-ről jö-t-tek volna le a színpad-ról
olyan jet lag-em van mint az állat
jött ez a pick-up truck-os nagyon mosoly-gó-s, bajusz-os ember
ott vaják-ol-t-ak, ott varázs-ol-t-ak, new age-el-t-ek
ki-hirdet-t-ék ez-t a Katalin Karikó appreciation day-t
a french fries-ra
nem french fries-nak fog-ják hív-ni
hanem át-nev-ez-ik freedom fries-ra
még mindig trash talk-ol
két interjú-ban is trash talk-ol-t rólam
én már csak side quest-ek-et csinál-ok
ez-zel a fajta small talk-kal ami-t mond-asz
ez persze semmi-t nem változ-tat a live stream-en
a gaming content-et hiányol-od?
most a main game-ek-et nem néz-ik annyian
én nem ugrándoz-om ez-ek-ért a fake meat-ek-ért