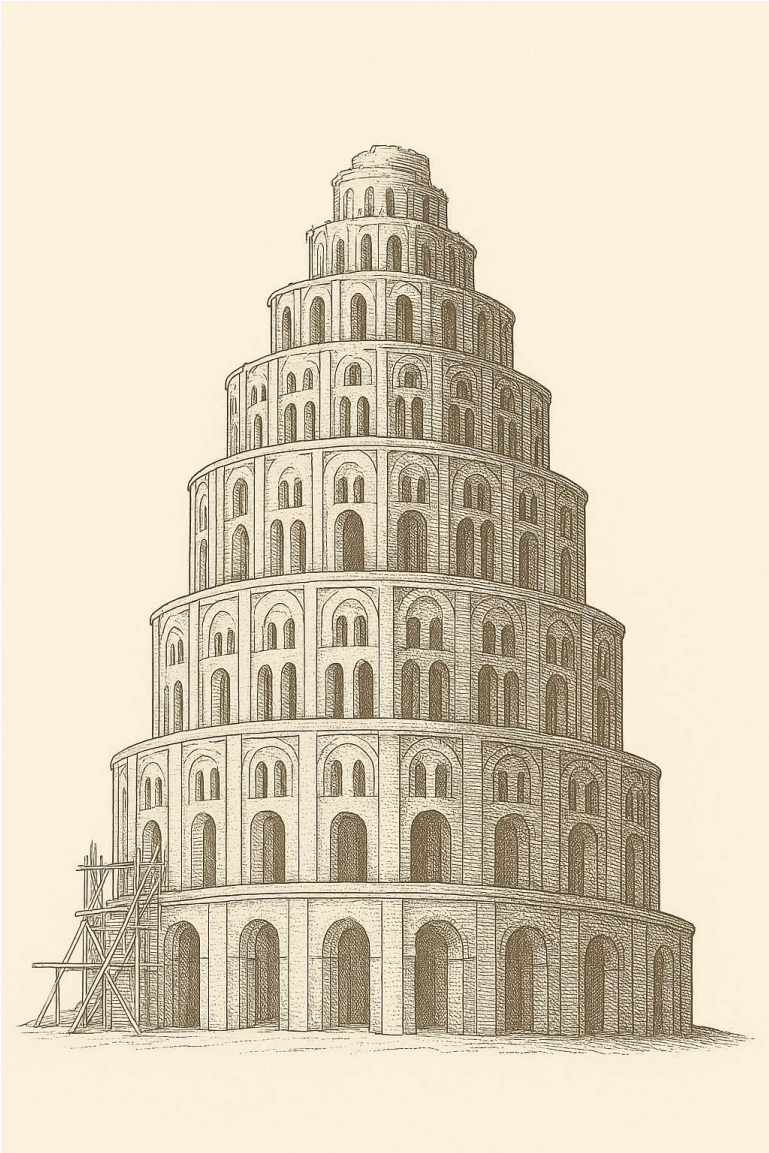
Polyglot Learning in the AI Era –

Design and Testing of an AI-Guided Simultaneous Learning Programme for the Acquisition of Related Languages ('die Unvollendete', version 8)



Dr. R. Schimmel

18 August 2026

Abstract

This dissertation presents research into the conditions under which three related Romance languages — Spanish, Italian and Portuguese — can be acquired simultaneously to CEFR A2 level within an accelerated learning programme, without explicit grammar instruction and with the support of AI tools. The research design combines design-oriented and autoethnographic research and is based on four interrelated assumptions: acceleration through pattern-rich input, positive transfer between related languages, metalinguistic reflection as a learning mechanism, and the emergent development of grammatical knowledge.

At the centre of the research is a learning design in which frequency-driven input, interlingual pattern comparison, controlled variation and scaffolding are combined. The theoretical principles have been operationalised in a thirteen-week learning programme consisting of an intensive receptive Week 0 followed by twelve weeks of gradual progression towards functional language proficiency. Semantically equivalent triplets in Spanish, Italian and Portuguese, micro-contexts, cloze exercises and interlingual strategies form the core of the learning materials. Generative AI is used for controlled material production and verification, while instruments for retention, pronunciation and reflection serve both instructional and research functions.

During the development of the learning materials, it became apparent that large-scale AI-driven production required additional design measures. Problems including semantic drift, loss of interlingual equivalence, insufficiently exclusive micro-contexts and cumulative variation led to a more tightly specified production architecture based on micro-batches, serial derivation, bounded contexts and explicit verification steps. The empirical language-acquisition test has yet to be conducted; the evidence requirements, analytical methods and interpretative framework for the three research questions have been specified in advance.

The contribution of the research therefore lies at two levels. For language acquisition, a testable design approach is developed in which structural relatedness is used not only as a potential source of interference, but also as a potential learning resource. With regard to AI-driven knowledge production, the research provides insight into how generative systems can be deployed within a process in which production, control, verification and human curation are functionally distinguished. The empirical phase must establish the extent to which this combination actually leads to accelerated simultaneous acquisition to CEFR A2 level.

Management Summary

Polyglot Language Acquisition in an AI Era

Introduction – a three-part research design. This study consists of three interconnected components: the dissertation, in which the theoretical, design-related and methodological foundations for AI-facilitated polyglot language acquisition are developed; the learning book, in which these principles are operationalised into a thirteen-week learning programme for the simultaneous acquisition of Spanish, Italian and Portuguese; and the Methodological Reflection Report, in which the role of artificial intelligence (AI) in knowledge production, material development and the creation of the study is reconstructed and critically accounted for.

This Management Summary presents the theoretical argument, the learning design, the research methodology, the empirical testing that remains to be conducted, and the interpretative framework already developed. The empirical phase has not yet been completed. The theoretical principles have been translated into a concrete learning design, the production architecture and learning materials have been developed, and the evidence requirements, analytical methods and structure for interpretation and conclusion have been specified in advance. The purpose of the empirical phase is to put this research model into practice and populate it with data.

The three components are functionally interconnected. In the dissertation, theory, design and research methodology are brought together; in the learning book, the design principles are translated into concrete learning activities and instruments for data collection; and in the Methodological Reflection Report, the use of AI in the production, control and iterative development of both text and learning materials is reconstructed. Together, they constitute a single research design in which language acquisition, instructional design, AI use and methodological control are aligned.

Chapter 1 – Introduction and positioning. Chapter 1 delineates the problem domain within which the study originated. The simultaneous learning of several related languages has received relatively little attention within conventional language pedagogy. Multilingual learning programmes are generally organised sequentially, whereas structural relatedness between languages may offer opportunities for transfer, pattern recognition and systematic comparison. At the same time, large language models make it possible to generate multilingual examples, variants, feedback and comparative material on a large scale. This gives rise to a new design question: not whether such systems can be used in language acquisition, but under what conditions they can contribute to accelerated multilingual acquisition without causing cognitive overload, interference or loss of grammatical understanding.

Against this background, the central research question is formulated as follows: *under what conditions can AI-facilitated support contribute to accelerating the simultaneous acquisition of*

three related languages to CEFR level A2, without loss of grammatical understanding or functional proficiency?

Four interrelated assumptions underlie this central question. First, rich, varied and immediately available input is assumed to facilitate pattern recognition and thereby accelerate language acquisition. Second, structural relatedness is regarded not only as a source of interference, but also as a potential instructional mechanism for positive transfer. Third, metalinguistic awareness is treated as a necessary condition for making productive use of similarities and differences. Fourth, grammatical knowledge is assumed not to require construction exclusively through explicit instruction, but also to be capable of emerging through repeated exposure, comparison, pattern recognition and directed reflection.

These assumptions are operationalised in three empirical subquestions: RQ1) To what extent can A2 level be achieved in three related languages within an accelerated, grammar-free learning programme supported by AI tools? RQ2) How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme? RQ3) To what extent does structural relatedness between the target languages function as an accelerator of simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?

Acceleration, transfer, reflection and emergence are thus treated as interrelated mechanisms: theoretically grounded in Chapter 2, operationalised in the learning design, and made separately and jointly testable in the empirical phase.

Chapter 2 – Theoretical framework. Chapter 2 develops the theoretical foundation of the learning design by combining theories of language acquisition, multilingualism research, cognitive mechanisms and insights from computational language modelling. The central question is under what conditions the simultaneous acquisition of three related languages is theoretically plausible without explicit grammar instruction.

The chapter first examines the specific opportunities and risks of learning in a multilingual context. Structural relatedness can facilitate positive transfer and faster pattern recognition, but simultaneously increases the risk of interference and cognitive overload. Against this background, various models of language acquisition are discussed, ranging from behaviourist and nativist approaches to communicative, interactional and, in particular, usage-based theories. The latter perspective is especially relevant to the design developed here: grammatical knowledge can gradually emerge from frequency, variation and recurring language use in meaningful contexts.

The chapter then examines the extent to which large language models can serve as a theoretical mirror for a learning environment centred on pattern recognition. No equivalence between human and computational language acquisition is implied. LLMs do not possess human meaning-making, intention or consciousness, but they do demonstrate that grammatically structured output can emerge from large-scale statistical exposure to language. For the learning design, their instructional potential is particularly relevant: generative models can provide large quantities of varied input, produce alternative formulations, support interlingual comparison and offer temporary scaffolding. At the same time, their limitations are explicitly taken into

account, including inconsistent output, pragmatic inaccuracy, cognitive overload and the black-box problem.

Five linguistic domains — phonology, morphology, syntax, semantics and pragmatics — are additionally used as an organising framework for similarities and differences between the target languages. From these domains, transfer, interference and metalinguistic awareness are developed further. Transfer is treated not solely as a source of error, but as a potential instructional mechanism. The interlingual strategies subsequently used in the learning design are theoretically grounded here as instruments for systematically recognising similarities, identifying divergences and supporting reflection on linguistic structure.

In the final part of the chapter, the theoretical framework is broadened to include language aptitude, linguistic complexity and structural distance. This serves not only to explain why simultaneous acquisition might be possible in general, but also why the specific combination of Spanish, Italian and Portuguese was selected for this experimental design. Cognitive load is related to four structural dimensions: phonological predictability, morphological transparency, syntactic stability and lexical relatedness. Across these dimensions, the three Romance languages display a relatively favourable combination of predictability, structural similarity and systematic divergence. This creates a learning environment in which relatedness is sufficiently strong to enable transfer, while differences remain sufficiently visible to make comparison and induction meaningful.

This theoretical synthesis leads to a design conclusion: AI, language relatedness and pattern-rich input do not in themselves constitute a learning process. Acceleration becomes plausible only when these elements are embedded in a controlled instructional environment in which variation, comparison, reflection and scaffolding are systematically aligned. Chapter 2 therefore prepares the design choices that are developed concretely in Chapter 3.

Chapter 3 – Learning design. Chapter 3 translates the theoretical principles from Chapter 2 into a concrete, executable and testable learning design. The underlying assumption is that grammatical knowledge need not be provided through explicit rule instruction, but can emerge from frequency, controlled variation, interlingual comparison and repeated exposure to meaningful language input.

The design philosophy is developed around the principle of *learning like an LLM*. This does not imply an equivalence between human and computational language acquisition, but refers to a distributional learning principle: structural regularities are inferred from large numbers of examples rather than taught in advance as grammatical rules. The fundamental distinction remains that human language learners can attach meaning, intention and reflection to those patterns. It is precisely this combination of pattern-rich exposure and conscious reflection that forms the core of the design.

The choice of Spanish, Italian and Portuguese is subsequently translated into an interlingual pedagogy in which Latin functions as a genealogical and organisational reference framework. Latin is not an independent learning target, but serves as a historical key through which recurring word stems, suffixes, sound correspondences and formal relationships between the three modern languages can be understood. This structural relatedness is operationalised through interlingual strategies and semantically equivalent triplets: parallel sentences in

Spanish, Italian and Portuguese in which a single conceptual core is made visible under controlled formal variation.

An explicit design grammar has been developed for the production of the learning materials. Variation is not generated freely, but restricted to predetermined grammatical dimensions and structured so that pattern recognition remains possible without cognitive overload. Cloze sentences function as instruments for active reconstruction. Particular use is made of micro-contexts that permit only one plausible lexical completion and must be realisable in all three target languages without semantic shift. Semantic equivalence is therefore treated not as an automatic outcome of translation, but as an explicit design requirement.

The temporal architecture consists of Week 0 followed by a twelve-week programme. In Week 0, 600 lemmas per language — 1,800 language forms in total — are presented receptively in order to activate phonological, lexical and morphological pattern recognition. Over the subsequent twelve weeks, the emphasis gradually shifts from receptive recognition towards structural, semantic and pragmatic application. Scaffolding is progressively reduced: external support through material structure, AI, repetition systems and reflection formats gradually gives way to internal guidance based on patterns already recognised.

Instrumentation is an integral part of the design. Generative AI is used for controlled material production and verification, Anki for spaced repetition, Speechling for the auditory component, and reflection formats for developing and recording metalinguistic awareness. The learning design therefore also functions as a research instrument: frequency, errors, retention, pronunciation and reflection generate distinct but combinable data streams through which the hypothesised learning mechanisms can be tested empirically.

During the actual development of the learning materials, it became clear that the original design principles alone were insufficient to keep large-scale production manageable. Problems such as semantic drift, insufficiently exclusive micro-contexts, formally correct but semantically meaningless sentences, loss of semantic equivalence between languages and the combinatorial growth of possible variants required further refinement. The production architecture was therefore reorganised around smaller production batches, strictly bounded contexts, serial derivation from a single reference language and controlled variation along one axis at a time. Variation in the derived languages is linked to the structure designed in the reference language, thereby preserving interlingual comparability.

Chapter 3 therefore ends not with a theoretical blueprint, but with a design refined through the process of material production itself. The principles of grammar-free, inductive and simultaneous polyglot learning have been retained, but embedded within a production and control system in which semantic stability, comparability and manageability are explicitly safeguarded. The learning book provides the operational implementation of this design and, at the same time, the instrument for the empirical testing that remains to be conducted.

Chapter 4 – Research methodology. Chapter 4 develops the methodology for empirically testing the learning design. An autoethnographic, design-oriented n=1 design has been adopted in which the researcher simultaneously acts as language learner, designer and documenter. The learning programme constitutes both the object of study and the environment in which data are generated. The limited generalisability resulting from this design is addressed through intensive

process documentation, pre-specified evidence criteria, triangulation and transparent documentation.

An important distinction is made between two types of data. On the one hand, **emergent learning processes** are recorded indirectly through behavioural and performance data, including retention, error patterns, response times and pronunciation development. On the other hand, **intentional data** are collected through logbook reflections, explanations of choices and other forms of metalinguistic self-observation. Combining these two layers of data makes it possible not only to establish whether performance changes, but also to investigate which learning mechanisms are associated with those changes. The learning book and the instruments incorporated into it therefore serve both instructional and data-collection functions.

For each of the three empirical subquestions, the type of evidence required, the data from which that evidence must be derived, and the alternative explanations that should be excluded as far as possible have been specified in advance. Functional proficiency is assessed using, among other things, CEFR-based tasks, retention data and pronunciation data; metalinguistic awareness is examined through reflections, pattern identification and application in new contexts; and language relatedness as an accelerating factor is analysed through speed, error frequency, transfer observations and contrasts between related and divergent structures.

Data analysis combines qualitative and quantitative approaches. Reflections and logbook data are coded thematically, while retention, error frequencies, task duration and other performance data are analysed quantitatively. Evidential strength is sought in convergence between different data streams: a hypothesised learning mechanism receives stronger support when, for example, improved task performance coincides with more stable retention and with reflections in which the same pattern is recognised and applied. Baseline measurements, novel task contexts, control tasks and fixed evidence standards are additionally used to distinguish memorisation, chance correctness and general growth in proficiency as far as possible from the mechanisms under investigation.

The quality of the AI-generated learning material is also treated as a methodological variable. A layered control procedure is therefore provided in which AI output is systematically checked for, among other things, grammatical correctness and semantic parallelism, followed where appropriate by external validation on a sample basis. The precise allocation of production and control functions between the AI systems used changed during development; methodologically, the central principle is therefore not any particular combination of tools, but the requirement that generated material should not be used as research input without prior control.

Finally, the dual role of researcher and language learner is explicitly treated both as a methodological limitation and as a source of data. Subjectivity is not assumed to be eliminable, but is channelled through fixed formats, pre-established coding principles, triangulation and version control of prompts, outputs, corrections and reflections. Before the empirical phase begins, it has therefore been specified which data will be collected, how they will be assessed and on what grounds claims regarding the three research questions may be made.

Chapter 5 – Empirical testing. Chapter 5 sets out the empirical reporting structure in advance on the basis of the evidence requirements and evaluation criteria established in Chapter 4. Because the learning programme has not yet been implemented, the chapter does not yet contain

research findings. It does, however, specify which empirical data are required to answer the three subquestions, how those data will be organised and according to which criteria a subquestion may be answered negatively, partially positively or fully positively.

For the first subquestion, functional proficiency at CEFR A2 level is linked to four separate evidence conditions: communicative performance, demonstrable progress in vocabulary and sentence production, external confirmation of intelligibility and self-assessment of functional language proficiency. For the second subquestion, explicit pattern recognition, application in new contexts, development of metalinguistic depth and growth in metalinguistic awareness are distinguished as separate components of evidence. For the third subquestion, speed and error reduction in related structures, explicit reflection on language relatedness and contrasts between related and divergent cases are used to establish whether structural relatedness has in fact functioned as an accelerating factor.

Chapter 5 thus functions as a preconstructed model for empirical reporting. The data still to be collected are assigned a fixed place within an evidence structure derived directly from the research questions and methodological criteria. This prevents the assessment criteria from being adjusted retrospectively to the findings.

A micro-iterative reporting structure has also been included for three planned reflection points during the learning programme. For each iterative phase, the empirical signals observed, their effect on the learning process or materials, the resulting design adjustment and the sources of evidence supporting that decision are recorded. In addition to the final outcome, this therefore documents how the learning design develops during implementation in response to empirical observations.

Chapter 6 – Conclusion, discussion and implications. Chapter 6 develops the interpretative and concluding framework within which the future empirical findings will be assessed. For each of the three subquestions, a fixed structure is provided in which the empirical answer is followed by interpretation in relation to the learning design and by discussion of validity, reliability and possible alternative explanations.

For the first subquestion, the emphasis lies on the feasibility of functional A2 proficiency within the accelerated grammar-free programme; for the second, on the relationship between metalinguistic awareness, pattern recognition and application; and for the third, on the extent to which structural relatedness has actually produced acceleration or error reduction. The separate findings are subsequently brought together in answering the central research question. The central issue is not the sheer quantity of AI-generated input, but the combination of controlled variation, repetition, reflection and human design direction.

An additional layer is provided by a predeveloped scenario framework in which a range of possible empirical outcomes has been considered systematically. The scenarios range from convergence between functional proficiency, metalinguistic awareness and transfer to partial patterns of success, language asymmetry, excessive dependence on AI, delayed convergence and a completely non-functional trajectory. Characteristic *red flags* and possible design implications have been identified for each scenario. Outcomes that only partially support, or directly contradict, the original expectations have therefore also been anticipated.

This establishes in advance which combinations of evidence can support different conclusions, which alternative explanations must be taken into account, and which findings may provide grounds for confirmation, adjustment or fundamental revision of the learning design.

Chapter 7 – Implications and concluding reflections. Chapter 7 places the developed research design within a broader theoretical and applied perspective. The transition to the empirical phase concerns the extent to which the pre-specified mechanisms and conditions hold in practice.

Several directions for further research are identified. Replication with larger groups of language learners is required to establish the robustness of the design principles beyond the autoethnographic n=1 context. Longer learning trajectories, external assessment and application to less closely related languages are also identified as relevant next steps. The present case thus functions as a bounded initial test environment for a broader research agenda on AI-facilitated polyglot language acquisition.

The theoretical and practical scope of the design is also delimited. The combination of semantically equivalent triplets, pattern recognition, interlingual comparison and scaffolding was developed specifically for accelerated functional acquisition to A2 level. At higher CEFR levels, additional requirements become relevant, including discourse competence, register sensitivity, extended interaction and the ability to handle abstract or specialist subjects. Scaling to B1–C2 is therefore likely to require an extension of the instructional architecture with different types of tasks and communicative situations.

A separate theoretical development concerns the possibility of an **interlingual grammar**. Existing grammars generally describe individual languages, while projects on intercomprehension[1] provide only partial systematisations of language relatedness. Section 7.5 therefore explores whether LLMs could be used to externalise and systematically organise patterns from parallel corpora of related languages. Such a family-wide meta-grammar does not yet exist, but could open a third position between explicit grammar instruction and entirely grammar-free learning: learning from an emergent rule space derived from systematic correspondences and divergences within a language family. AI could thereby potentially serve not only as an instructional tool, but also as an instrument of knowledge production within comparative linguistics.

The contribution delineated in Chapter 7 therefore lies in the combination of theory, learning design and methodological instrumentation. The result to date is not demonstrated acceleration, but a systematically developed and empirically testable research design. Its limitations — n=1, an empirical phase that has yet to be conducted, and dependence on generative technology — remain part of the interpretation of future findings.

[1] **Intercomprehension:** the ability to understand related languages without fully mastering each of them separately, by systematically drawing on shared vocabulary and structural similarities.

Learning book and Methodological Reflection Report. The learning book forms the operational counterpart of the dissertation. It translates the theoretical principles into an executable thirteen-week learning programme, with Week 0 as an intensive receptive starting phase followed by twelve weeks of gradual progression towards functional proficiency at A2 level. The interlingual strategies, semantically equivalent triplets, micro-contexts, controlled variation, scaffolding and reflection formats take concrete instructional form within it. At the same time, the learning book is structured as a research instrument: exercises, retention data, production tasks and reflections provide the data through which the central assumptions can be tested empirically.

The development of the learning book gradually acquired a second methodological significance. It was precisely the production of large quantities of tightly specified multilingual learning material that exposed limitations easily concealed in ordinary text generation. Semantic drift, loss of interlingual equivalence, insufficiently exclusive micro-contexts, cumulative divergence and the combinatorial growth of variation possibilities required an increasingly formalised production process. The learning book thereby also became a stress test of AI-driven production: an initially prompt-driven process evolved into a production architecture based on micro-batches, serial derivation, explicit stop rules, separate production and verification steps, and an increasingly precise division between human and artificial tasks.

The **Methodological Reflection Report** reconstructs and critically accounts for this development. Its function extends beyond a conventional statement on AI use. It records that textual production of the dissertation was fully delegated to generative AI, while problem formulation, selection, process architecture, assessment, verification and scientific responsibility remained with the human curator. Successive phases in the development of the dissertation and learning book are linked to the underlying interactions, allowing not only the final product but also the process by which it came into being to be examined.

This reconstruction shows how AI was successively used for domain exploration, theoretical synthesis, structuring, text production, revision, material production, critique and verification. The use of different models proved not to be tied to fixed intrinsic roles: production and critique could be distributed between systems and, in other projects, could also be reversed. Control was therefore organised not around trust in a single model, but around a combination of production, contradiction, verification and human judgement.

Control of scientific source use is also developed in the Reflection Report as a separate methodological layer. Bibliographic identity, local claim–source use and the transition from existing knowledge to new theoretical or design-oriented synthesis are controlled separately. This exposes a characteristic of AI-driven knowledge production that cannot be addressed through bibliographic verification alone: substantively plausible synthesis can coexist with unstable provenance or overly broad claim–source linkages. Control therefore concerns both whether sources exist and where existing knowledge ends, where legitimate theoretical extension begins, and where a new construction might incorrectly be presented as established knowledge.

The Reflection Report thus has a dual function. For the language-acquisition study, it provides a methodological account of how the dissertation and learning instrument were produced. Within the broader study of AI-driven knowledge production, that developmental history itself

also constitutes empirical material: it demonstrates how full delegation of textual production does not eliminate human scientific work, but relocates it towards problem formulation, curation, process design, verification and responsibility.

Conclusion regarding the three-part research design. At its current stage of development, the study is conceptually and methodologically advanced, while the empirical language-acquisition test remains to be conducted. The theoretical positioning, learning design, production architecture, evidence structure, instrumentation and interpretative framework have been developed; several possible patterns of future outcomes have also been distinguished in advance. The empirical phase must establish which of the hypothesised learning mechanisms actually occur and whether the intended acceleration is achieved. The *Unvollendete* has thus acquired its form, while the empirical outcome remains open.

The substantive contribution lies in bringing together four elements that are usually treated separately: simultaneous acquisition of related languages, emergent learning without explicit grammar instruction, metalinguistic reflection, and AI-supported production and scaffolding. This has produced a research design in which relatedness is treated not only as a source of interference, but also as a potential learning resource, and in which language input is organised as a controlled pattern space. Empirical implementation will have to establish the extent to which this combination produces the intended effect.

The development process has also produced insights that do not depend on the ultimate outcome of the language-acquisition test. In producing the dissertation and learning book, a shift became visible from simple prompt steering to process and production architecture; from trust in coherent output to explicit verification; from a single generative system to organised counterplay between different systems; and from human text production to human curation. This also made clear that the quality of AI-generated academic production depends not only on the capabilities of the language model used, but also on how generation, critique, verification and human decision-making are organisationally separated.

In retrospect, the study therefore acquired a role that had not been anticipated at its inception. It originated before the pentalogy *AI-driven Knowledge Production,* but became its historical first case and represents the design-oriented archetype within it. Several practices that could be specified in advance in the later four studies still had to be developed during the process here: the curatorial role of the human researcher, the use of AI as a text producer rather than solely as a writing assistant, the interchangeability of production and critique roles between models, adversarial control, verification of sources and claims, and the transition from prompting to reproducible process architecture. These experiences yielded methodological principles that could subsequently be specified and applied more systematically in the later studies.

Once the pentalogy has been completed, this first case acquires renewed significance. What still had to be developed during the process here can then be understood as part of a broader methodological framework for AI-driven knowledge production. The question thereby shifts from whether full AI text production under human curation is possible to what this division of labour means when it is systematically applied across diverse scientific and societal domains of knowledge. The *Unvollendete* thus stands chronologically at the beginning of the pentalogy and has also provided important methodological principles for what followed.

Table of Contents

Prologue – The Three Creations

"Each and every language maps a possible world, a possible calendar and landscape."
George Steiner, *The Poetry of Thought*

There was a time when the world spoke without words. Sound, scent and gesture were enough for survival; meaning remained bound to the immediate. Then human beings learned that a sound could evoke something that was absent. With that discovery began humanity's first creation of its own: language.

1. The first transition – the biological revolution. From the convergence of three capacities — motor, cognitive and social — emerged a self-reinforcing system ('autocatalytic system'): the capacity to produce language. Control over the speech organs made the differentiation of sounds possible; the capacity for representation gave those sounds content; and shared attention gave them direction. *Language made knowledge transmissible without heredity: learning replaced mutation as the vehicle of evolution. Humanity broke free from the pace of nature.* Where other species adapted, human beings began to reshape their environment. Language became the engine of cooperation, planning and imagination — the catalyst of culture. In the breath with which humanity named the world, it began to shape it: human beings took the helm of creation.

2. The second transition – the media revolution of writing. Centuries later, humanity found a way to preserve its voice. Writing extended memory, but also changed the nature of language itself. What had been fluid became fixed. What had been shared became property. *Writing made knowledge durable, but also established boundaries.*

This also gives meaning to Steiner's image of every language as a possible world, a possible 'calendar' and a possible 'landscape'. Language not only orders what can be named, but also helps determine how time, space and experience are distinguished. Different languages therefore do not necessarily map the same reality in precisely the same way. Through writing, such linguistic orderings became durable and could become more sharply differentiated from one another.

The natural variation of dialects was codified, and codification hardened that variation into separate languages. The myth of Babel describes this solidification as the confusion of tongues: humanity, seeking to preserve its language, made it less intelligible to others. Unity became division: not through a loss of words, but through too much order. Language, once a continuum of sound and meaning, became an archive of rules. Writing extended the voice, but tamed it; it gave humanity power over knowledge, but deprived language of its fluidity.

Yet this process also gave rise to a counter-movement: ***polyglottism.*** Those who learn several languages once again hear the resonance of a single voice beneath the many. Polyglottism is the ability to listen through codification, to recognise the relatedness that writing once concealed. It is an attempt to restore natural mutual intelligibility — a human correction to the Babylonian fixation.

3. The third transition – the cognitive revolution of AI. Now, millennia later, language once again steps beyond the human body. Artificial intelligence uses language not to speak, but to analyse. It does not understand, but orders; it has no intention, but recognises patterns. In its algorithms, humanity sees itself reflected once again: a third creation in which language is no longer a means, but a mirror.

In biblical terms, AI is the *tree of the knowledge of good and evil*: it tempts through the promise of all-encompassing knowledge, it is dangerous because it strips the world of mystery, and it is illuminating because it reveals how that mystery was once created through language itself. In the spirit of Max Weber, AI marks a new phase of *disenchantment*: the reduction of meaning to calculation, of insight to probability. Yet precisely through that loss of magic, wonder may return — not as belief, but as insight into the structure of knowledge. AI is therefore both a threat and a *blessing in disguise*: it shows how what diverged through evolution and was hardened by writing can be reconnected. In its digital space, patterns, words and meanings flow together once more: the divergence perpetuated by history is temporarily suspended through algorithmic ordering.

4. Bridge to this research. This dissertation is situated within that tension — between separation and reconnection, between loss and insight. It examines how AI, at once a risk and an instrument, can reactivate the human capacity for language. Not to replace human beings, but to remind them of what language fundamentally is: a system of sounds, patterns and intentions continually seeking coherence. In the study of three related languages — Spanish, Italian and Portuguese — this third creation becomes tangible. Polyglottism is understood here not as a trick, but as a form of symbolic restoration: an attempt to experience lost unity once again through artificial ordering. What follows begins where this prologue ends — with the question of how human beings, assisted by a machine, can learn once again to hear what connects all languages[2].

Remco Schimmel,

18 August 2026

[2] The observations in this prologue are inspired by insights from theories of language evolution (Bickerton, 1990; Deacon, 1997; Hauser, Chomsky & Fitch, 2002; Tomasello, 2008), by the account of writing as a media rupture in the history of communication (Ong, 1982), and by Weber's (1919/2002) sociological-philosophical interpretation of rationalisation and disenchantment. The metaphor of the Babylonian confusion of tongues draws on linguistic analyses of language divergence and standardisation in the Romance context, as discussed by Posner (1996) and Maiden (2011). The interpretation of polyglottism as a restorative counterforce is also informed by modern approaches to metalinguistic awareness and multilingual cognitive representation (Bialystok, 2001; Kroll & Bialystok, 2013). The opening quotation and its conception of languages as different orderings of world, time and space are taken from George Steiner, *The Poetry of Thought* (2011).

Chapter 1 – Introduction and Positioning

§1.1 Exploring the Problem Domain

In international education, the need for multilingualism is increasing, while practical reality continues to lag behind. Most pupils and students acquire no more than one foreign language to a functional level, often with considerable difficulty, over a prolonged period, and with little awareness of transferable strategies. Multilingualism is encouraged at policy level, but is rarely facilitated through instructional design.

At the same time, the technological infrastructure of language acquisition is changing fundamentally. The emergence of large language models (LLMs), such as OpenAI's ChatGPT and Anthropic's Claude, makes it possible to generate example sentences in natural language, request error correction, explain differences in meaning and even automate multilingual comparison. For the first time in history, a language learner can independently — without a teacher, textbook or classroom — generate virtually unlimited input, feedback and variation in several languages simultaneously.

Whether this actually leads to better or faster language acquisition, however, remains far from established. The point of departure is therefore not a technological promise, but a testable assumption: suppose an LLM has the potential to accelerate language acquisition — under what conditions would this be the case? And would the same apply to learning several related languages simultaneously?

This brings a second premise to the fore: multilingual language acquisition need not necessarily proceed sequentially. Learning related languages simultaneously — Spanish, Italian and Portuguese, for example — might instead produce cognitive synergy. Their overlap in vocabulary, grammar and sentence structure offers opportunities for structural transfer, pattern recognition and reflection.

The focus of this dissertation is therefore not whether LLMs can be used in language education — this is already happening worldwide — but how such an environment must be designed in order to facilitate multilingual language acquisition without causing cognitive overload, loss of grammatical depth or instructional dilution.

§1.2 Polyglottism: More Than Three Times One Language

Polyglottism is often defined as the ability to speak several languages. In the context of this research, however, it involves more than the accumulation of languages. A learner acquiring three related languages simultaneously does not simply develop three separate language competences, but builds an internal grammatical space: a mental system in which relatedness, contrast and function are systematically organised.

A polyglot learns not only that something differs between language A and language B, but also why that difference matters — pragmatically, structurally or culturally. As Jessner (2006) and

Cenoz and Gorter (2011) argue, this capacity for comparative language use is not merely an end product of multilingual acquisition, but also functions as a means of learning: the process of structural comparison stimulates language awareness and metalinguistic development. Through systematic comparison — What changes if I say this in Spanish? How would this sound in Portuguese? — not only multilingualism but also a metacognitive layer emerges: language awareness.

Polyglot language acquisition therefore differs fundamentally from sequential, one-language-at-a-time learning. It presupposes active interlingual reflection, requires a learning environment in which differences are visible, identifiable and meaningful, and demands a certain degree of self-direction: the language learner must gradually become able to structure personal observations, interpret errors and recognise patterns across language systems.

Research on multilingual acquisition suggests that polyglots generally display stronger metalinguistic reflection, pattern recognition and flexible transfer (Jessner, 2006; Bialystok, 2001). At the same time, beginners are known to struggle with abstract grammatical concepts and to be misled by superficial similarities — the so-called false-friends effect (for example, Spanish *embarazada* means 'pregnant', not *embarrassed*). This tension forms the instructional core of the dissertation: how can a learning environment be designed in which polyglottism does not produce confusion, but instead accelerates learning? And how can AI function as an instrumental factor within such an environment?

§1.3 The Choice of Spanish, Italian and Portuguese

According to Posner (1996) and Maiden and Robustelli (2007), these languages share a common origin in Vulgar Latin, with substantial overlap in core vocabulary and morphosyntactic structure (i.e. the ways in which words are inflected and combined into sentences). At the same time, contrastive analyses (Rothman, 2009) show that the three languages differ sufficiently in pragmatics, such as forms of politeness, phonology and word order to cause significant interference when learned simultaneously. This combination of relatedness and variation is instructionally interesting. On the one hand, it enables the language learner to activate three systems simultaneously with relatively few entirely new learning elements. On the other hand, it demands precision: resemblance does not automatically imply equivalence.

The three languages are also socially relevant. Spanish is spoken worldwide by more than 500 million people across Europe, Latin America and the United States. Portuguese is the official language of Portugal, Brazil and parts of Africa, with approximately 250 million speakers in total. Italian is not only a major language of culture, but also an important language in fashion, design, gastronomy and music.

For language learners active in Europe, South America or international professional contexts, these languages together form a powerful combination. Precisely because of their relatedness, proficiency in all three may offer a strategic advantage. Research on multilingual receptive skills (Dewaele & Wei, 2013; Rothman, 2009) indicates that related languages can benefit from enhanced opportunities for transfer, faster lexical recognition and improved access to grammatical structures. It has also been suggested that this can lead to more efficient functional

language use in situations in which several Romance languages are used alongside or in combination with one another (cf. Grosjean, 2010).

It is therefore assumed that this particular combination of three related languages is especially suitable for simultaneous acquisition — provided that the learning environment is designed to exploit relatedness as structure rather than treating it as noise. The underlying assumption is that grammar-free, AI-driven polyglot acquisition is possible provided that the learning design safeguards semantic consistency and controlled variation.

§1.4 Metalinguistic Reflection as an Accelerator of Language Acquisition

One of the central premises of this dissertation is that *metalinguistic reflection* — thinking about language structures, meanings and form–function relationships — can be a powerful learning mechanism, particularly in the simultaneous acquisition of related languages.

Metalinguistic awareness is defined in the linguistic literature as the ability to step back from the immediate communicative function of language and approach language as an object of analysis (Jessner, 2006; Bialystok, 2001). Learners who develop this ability acquire not only an understanding of *what* something means, but also *why* it is grammatically correct or appropriate — and can apply that understanding across language boundaries.

In the context of *polyglot language acquisition*, metalinguistic reflection is assumed to be not merely a by-product of learning, but a necessary accelerator. Someone learning three languages simultaneously must, in all likelihood, be able to recognise similarities and differences systematically, reflect on *why* something works differently in language A than in language B, and not merely correct errors but understand them functionally. The hypothesis underlying this research is that language acquisition can be accelerated not only through technological tools, but above all by training reflective abilities that make language a *conscious object of structuring* — without formally teaching that structure.

§1.5 Functional Learning Objectives within the CEFR

To enable language proficiency to be measured, compared and developed across different contexts, Europe has used the Common European Framework of Reference for Languages (CEFR) since 2001. The framework distinguishes six levels of language proficiency, organised into three broad categories:

- Basic user: A1 and A2
- Independent user: B1 and B2
- Proficient user: C1 and C2

This research focuses on level A2. According to the CEFR definition, A2 means that a person is able to understand sentences and frequently used expressions relating to personal and everyday topics, such as family, shopping and work; carry out simple communicative tasks,

such as introducing themselves or asking for something; and provide short descriptions of their immediate environment, habits and immediate needs.

The hypothesis in this dissertation is that A2 constitutes a critical threshold level. It is low enough to be achievable within a limited learning programme, even in three languages simultaneously. It is high enough to require functional language proficiency: the language learner must be able to speak actively, understand language and interpret short texts. It is also the level at which gains from acceleration — for example in vocabulary growth, sentence production or receptive recognition — first become concretely visible and measurable, without the complexity of the language input or grammatical variation increasing to such an extent that learnability is compromised.

As the CEFR (2020) makes explicit, A2 marks the point at which language use shifts from elementary formulae and isolated words (A1) towards independent functional communication in routine everyday situations. Lightbown and Spada (2013) likewise emphasise that this level is the first at which visible progress emerges in active skills, while Bialystok (2001) associates A2-like performance with the beginnings of controlled production and reflective language use.

A2 therefore constitutes a critical transition point: high enough to demonstrate functional skills, but low enough to keep interlingual comparison and error analysis manageable (cf. CEFR, 2020; Lightbown & Spada, 2013; Bialystok, 2001).

§1.6 Research Objective, Assumptions and Scope

This dissertation examines a bold assumption: that it is possible to acquire three related languages — Spanish, Italian and Portuguese — simultaneously to a functional level of proficiency (CEFR A2), without explicit grammar instruction and within a twelve-week learning programme. Not through conventional instruction, but through interaction with AI systems that generate examples, provide feedback and stimulate pattern recognition.

The issue is therefore not only whether this is possible, but also how it works, what it requires of the language learner, and which mechanisms underlie such acceleration. At the heart of the design lies the assumption that language acquisition is not a formula, but an interactive, reflective and transfer-sensitive process.

The **central research question** is therefore: *"Under what conditions can AI-facilitated support contribute to accelerating the simultaneous acquisition of three related languages to CEFR level A2, without loss of grammatical understanding or functional proficiency?"*

The assumption embedded in the central research question is based on four underlying assumptions, each addressing a different aspect of the learning process:

(I) Assumption regarding acceleration (AI): LLMs enable accelerated language acquisition through immediate input, feedback and variation. This assumption is consistent with insights from computer-assisted language learning (CALL) and intelligent tutoring systems (Chapelle, 2003; Colpaert, 2020), in which immediate feedback and contextual variation are regarded as factors that can accelerate acquisition.

(II) Assumption regarding transfer (language relatedness): Related languages can reinforce one another structurally, provided that explicit attention is paid to their differences within the learning process. This assumption builds on contrastive approaches to language acquisition (Odlin, 1989) and more recent work on cross-linguistic transfer between related languages (Rothman, 2015).

(III) Assumption regarding reflection (awareness): Language reflection is indispensable if relatedness is to function productively rather than become a source of confusion. According to Jessner (2006) and Cenoz (2013), metalinguistic awareness in multilingual speakers is a key factor in successful interlingual coordination and interference control.

(IV) Assumption regarding emergence: Grammatical understanding need not be explicitly taught; it can emerge through pattern recognition, repetition and purposeful interaction with AI.

These assumptions have led to the following **empirical subquestions:**

(1) To what extent can A2 level be achieved in three related languages within an accelerated, grammar-free learning programme supported by AI tools?

(2) How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme?

(3) To what extent does structural relatedness between the target languages function as an accelerator of simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?

These questions are addressed in Chapters 4 to 6. The analysis focuses systematically on what happens when a language learner studies three related languages simultaneously, without explicit rules but with targeted support — and whether this produces the outcomes anticipated by the design.

The next chapter establishes the theoretical foundation for this design approach. Among other things, it considers how LLMs relate to existing instructional models and how multilingual acquisition without explicit grammar instruction can be theoretically grounded.

Chapter 2 – Theoretical Framework

§2.1 Learning in a Multilingual Context: Principles and Challenges

Acquiring several languages simultaneously differs fundamentally from learning a single foreign language sequentially. Earlier studies of the simultaneous acquisition of related languages (Cenoz, 2000; Herdina & Jessner, 2002) indicate that learners studying three languages in parallel may face cognitive overload — that is, the processing of several linguistic systems at the same time places substantial demands on working memory — an increased risk of interference, and greater instructional complexity. These tensions are particularly apparent when the languages are structurally similar but functionally different. This chapter maps those tensions and outlines the theoretical framework within which simultaneous language acquisition can be examined.

A first premise is that related languages represent both an opportunity and a risk. Shared vocabulary, grammatical structures and phonological similarities can lead to positive transfer: recognising and applying comparable patterns in a new language can facilitate acquisition (Odlin, 1989; Ringbom, 2007). At the same time, there is a risk of interference or negative transfer, when a linguistic pattern from language A is incorrectly applied in language B, leading to systematic errors (Jarvis & Pavlenko, 2008). Successful polyglot learning therefore requires a guided learning process in which structural relatedness is made explicit and actively contrasted in order to avoid confusion.

A second premise is that simultaneous language acquisition — particularly in the case of structurally related languages — involves increased cognitive load, because several linguistic systems must be activated, coordinated and distinguished in parallel (De Bot & Jaensch, 2015; Gibson, 1998): Is this correct in all three languages? Is this form unique to one language? Have I made this error before in another language?

This complexity requires not only greater memory and attentional resources, but also a heightened degree of conscious language use: a metalinguistic orientation that is often not yet self-evident for beginners.

A third premise is the shift from language acquisition as rule-based learning towards language acquisition as pattern recognition. From this perspective, it is relevant to examine what happens when a language learner — like a language model such as ChatGPT — is exposed to large quantities of input (‘mere exposure’) without explicit grammar instruction. One of the assumptions explored in this research is that if humans learn in ways analogous to LLMs — through patterns, variation and exposure alone — this may lead to accelerated acquisition of functional language proficiency.

Set against this assumption is conventional language pedagogy, in which instruction, repetition and explicit explanation occupy a central role. The question is not whether AI can support multilingual acquisition, but how such technology should be embedded within a learning environment that provides scope for reflection, structuring and awareness of error.

Finally, it is important to recognise that the simultaneous acquisition of related languages is not a self-evident process. First, it requires a carefully designed learning environment in which

structural relatedness is deliberately used as an instructional resource rather than treated as an implicit assumption (cf. Cenoz & Gorter, 2011). Second, research on self-regulated learning and metacognition indicates that language learners need support in planning, monitoring and adjusting their multilingual learning process (Boekaerts & Corno, 2005). Third, simultaneous acquisition places distinctive demands on the way meaning, grammar and context are connected across several languages, an issue that is particularly relevant when designing tasks in which multiple languages are brought together functionally (Mehisto, 2012; García & Wei, 2014).

The following sections therefore examine, in turn, existing models of language acquisition, the role of LLMs, the linguistic domains that can serve as an organising framework, and the mechanisms of transfer and reflection that are central to this research.

§2.2 Models of Language Acquisition: From Behaviourism to Usage-Based Learning

To provide a meaningful foundation for the assumption of accelerated multilingual language acquisition through AI, an understanding is required of existing theories of how people learn languages. Four dominant approaches are discussed in this section, each with different implications for multilingual learning, self-direction and the use of digital tools.

1. Behaviourism: learning through imitation and reinforcement. In the early twentieth century, language acquisition was often understood within a behaviourist framework (Skinner, 1957). According to this model, learning results from conditioning: correct responses are rewarded and errors discouraged. Language learning is therefore a process of stimulus–response and repetition. The language learner assumes a largely passive role. For the simultaneous learning of related languages, this model is of limited use because it allows little scope for variation or interference. It does, however, emphasise repeated exposure and repetition — a principle that remains relevant in contemporary language pedagogy and digital learning tools.

2. Nativism: language as an innate capacity. Chomsky (1965) criticised behaviourism by arguing that human beings possess a universal grammar: an innate cognitive system that enables complex grammatical structures to be inferred from limited input. This provides an explanation for how children acquire language without explicit rules. For polyglot acquisition, however, nativism offers little practical guidance. It provides an explanation for the capacity to learn several languages, but not for the instructional design of such a process. Little attention is given to differentiation, feedback or the management of language interference.

3. Communicative and interactional approaches. From the 1980s onwards, attention shifted towards language as communication (Canale & Swain, 1980; Long, 1996). From this perspective, language is acquired through meaningful interaction, in which both input — comprehensible and rich — and output — producing language oneself — play central roles. Swain (1985) introduced the concept of pushed output: the idea that language acquisition results not only from listening and reading, but also from the need to produce comprehensible and accurate spoken or written language. Digital tools can play a role in this process. When a system offers an alternative or improved formulation, the learner is encouraged to reconsider their own

production. Research on intelligent tutoring systems and chatbots (Yilmaz & Granena, 2019) indicates that such feedback can contribute to syntactic development and conscious reflection on linguistic form.

4. Usage-based learning and pattern recognition. In more recent work, language acquisition is understood as the gradual development of patterns from usage (Ellis, 2003; Tomasello, 2009). From this perspective, linguistic structures are not acquired primarily through explicit rules, but through recurring usage events: concrete instances of language use in context. Frequency and function are central.

This closely corresponds to the idea that statistical patterns — rather than grammatical instruction — constitute the driving force of learning. Digital systems can provide support by generating rich variation in input, thereby making recurring patterns more visible and recognisable to the language learner.

§2.3 LLMs as a Mirror of Language Acquisition? On 'Learning without Grammar'

In recent years, so-called large language models (LLMs), also referred to as Generative Pre-trained Transformers (GPTs) — statistical models trained on very large text corpora — have attracted worldwide attention because of their ability to generate, interpret and edit coherent text. Well-known applications include ChatGPT, developed by OpenAI, and Claude, developed by Anthropic.

ChatGPT is an example of an interactive application in which users can produce, modify and analyse natural language with the aid of AI. ChatGPT is a specific type of LLM trained on hundreds of billions of words drawn from diverse text corpora, without being explicitly provided with grammatical rules or semantic concepts. Instead, the model 'learns' through pattern recognition: on the basis of statistical relationships, it predicts the most probable next word in a given context.

This LLM architecture was elaborated, among others, by Brown et al. (2020) in their description of GPT-3, one of the first large-scale models capable of zero-shot and few-shot language processing at human-like levels ('zero-shot' and 'few-shot' refer to the number of examples required to perform a new task, ed.). Although an LLM does not understand language in the way humans do — it has no meaning, intention or world knowledge — it can produce grammatically correct, contextually appropriate and communicatively functional sentences in dozens of languages. This raises an intriguing question:

What can the way an LLM processes language teach us about how humans might learn language?

A hypothetical comparison: learning as pattern recognition. In traditional language pedagogy, grammar often occupies a central place as explicit learning content. An LLM, however, shows that complex linguistic structures can emerge from exposure to rich and varied examples without explicit instruction. This closely corresponds to usage-based approaches to

human language acquisition (Ellis, 2003; Tomasello, 2009), in which pattern recognition and frequent contextual input are treated as the driving forces of learning.

The assumption explored in this dissertation is that when humans learn in an environment resembling that of an LLM — rich in variation, frequency and repeated exposure — they may be able to acquire linguistic structures more rapidly, even without explicit grammar instruction.

Mere exposure: sufficient? Or necessary but insufficient? The idea that language acquisition can occur through exposure alone (Zajonc, 1968) — often referred to as 'mere exposure' — has also been advanced in usage-based approaches, in which frequent, context-bound input provides the basis for pattern recognition and the development of constructions (Ellis, 2002; Tomasello, 2003). This can be observed in natural first-language acquisition: infants acquire language without explicit grammar instruction, through rich and repeated input in meaningful contexts. Adults can undergo a comparable process to some extent through immersion in a target-language environment, although this is generally less effective without guidance (Lightbown & Spada, 2013).

LLMs represent a radically scaled-up model of exposure: they process billions of sentences from highly diverse contexts. Unlike humans, however, they lack the capacity for conscious reflection — precisely the capacity in which human language learners may excel, provided that it is explicitly supported. This may mark the crucial distinction: LLMs process patterns but do not understand them; humans can ultimately understand patterns, provided that they learn both to recognise and interpret them. The question, therefore, is not whether humans should learn 'like an LLM' — that is impossible — but whether LLM-like learning environments, offering large numbers of examples, alternatives, variants and feedback, provide an instructional advantage for human language learners who learn through induction and reflection.

Definition and positioning of LLMs in this research. In this dissertation, LLMs are defined as follows: "A statistical language model that independently generates language through pattern prediction on the basis of very large text collections, without relying on explicit linguistic rules." An LLM is not treated here as an authority or as a replacement for human learning, but as an instructional tool: a source of examples, feedback and contrasts. It is this role — rather than the model itself — that is central to the research.

Critical reflection. An LLM is a so-called black box: it generates linguistic output without transparency regarding the precise operation of the underlying calculations and decision processes. For scientific use, this is problematic because reproducibility and verifiability are core requirements of empirical research. A black box may nevertheless be temporarily useful, provided that the behaviour of the system proves sufficiently consistent and reproducible within a bounded context. This does not remove the need for retrospective accountability: researchers must make explicit why the use of a black box was considered acceptable in the particular case, under what conditions its outputs were regarded as reliable, and where the limits of that reliability lie. Arbitrary behaviour is incompatible with scientific testing. Earlier research on AI-facilitated language education indicates that such systems can contribute functionally to enriched input, form-focused feedback and formative practice situations (Li, Link & Hegelheimer, 2015; Wang & Vasquez, 2012), provided that their limitations as non-learning systems are adequately compensated for through deliberate human guidance.

§2.4 Linguistic Domains as Cognitive Organising Principles

When several languages are learned simultaneously, sounds, words and structures can quickly become difficult to disentangle. To make this complexity manageable, a systematic organisation of linguistic phenomena is essential. This dissertation uses five classical linguistic domains that together describe the principal components of language structure and use. **Phonology** concerns the sound system of a language: how sounds are organised, varied and pronounced. **Morphology** describes how words are constructed from smaller meaningful units (morphemes) and how their forms change. **Syntax** concerns the rules by which words are combined into sentences and phrases. **Semantics** concerns the meaning of words and sentences independently of context. **Pragmatics** addresses language use in social context, including implicit meanings, politeness strategies and speech acts.

These domains are not treated here as abstract analytical categories, but as cognitive organising models that can help language learners recognise patterns and make differences manageable. Digital tools such as LLMs can support this process, for example by generating examples or context-specific variants that invite comparison and reflection. Organising language along these domains can help learners explain errors, recognise nuances and structure language intuitively, without explicit grammatical explanation always being necessary. This is consistent with findings from language-acquisition research in which conscious reflection on linguistic structure is regarded as a potential accelerator of learning (Lightbown & Spada, 2013).

Phonology – Sound and pronunciation. Phonology concerns the sound structure of a language: sounds (phonemes), stress, intonation, rhythm and pronunciation patterns. Although the three target languages are related, they display subtle but meaning-relevant phonological differences, including nasalisation, vowel variation and stress.

Phonological feature	Dutch	Spanish	Italian	Portuguese
Nasalisation	None	None	None	Present: *mão, pão*
Palatal nasal [ɲ]	None	*niño* [ˈniɲo]	*gnocchi* [ˈɲɔkki]	*senhor* [seˈɲoɾ]
Stress position	Variable	Usually penultimate	Usually penultimate	Highly variable
[u] sound	*boek* [buk]	*luz* [luθ / lus]	*luna* [ˈluːna]	*luz* [luʃ]

Systematic comparison of pronunciation promotes phonological awareness, which is essential for listening comprehension and intelligibility (Field, 2005).

Morphology – Word structure and inflection. Morphology concerns the ways in which words change form to express grammatical functions: verb conjugation, plural formation, gender, and so forth. In the Romance languages, verb morphology is particularly rich and conveys substantial information about tense, person and aspect.

Person / Tense	Dutch	Spanish	Italian	Portuguese
I speak	*spreek*	*hablo*	*parlo*	*falo*
You speak	*spreekt*	*hablas*	*parli*	*falas*
We spoke	*spraken*	*hablamos*	*parlavamo*	*falámos*
They have spoken	*hebben gesproken*	*han hablado*	*hanno parlato*	*têm falado*

Similarities in stem structure make morphological awareness possible, while differences in form invite reflection. Comparison strengthens learners' inductive capacity and their ability to reconstruct grammatical systems implicitly (Ellis, 2003).

Syntax – Sentence structure and word order. Syntax concerns the ordering of words within sentences. It determines how meaning is constructed through order, agreement and constituent structure. There are substantial differences between Dutch and the three target languages, particularly in the placement of so-called clitics — short word forms such as *me*, *te* or *lo* in Spanish — and in the structure of questions and negation.

Meaning	Dutch	Spanish	Italian	Portuguese
I see him tomorrow	*Ik zie hem morgen*	*Lo veo mañana*	*Lo vedo domani*	*Eu o vejo amanhã*
Tomorrow I see him	*Morgen zie ik hem*	*Mañana lo veo*	*Domani lo vedo*	*Amanhã eu o vejo*
Interference error	*Ik hem zie morgen*	*Veo lo mañana*	*Vedo lo domani*	*Vejo o amanhã*

Syntactic differences often become visible only through explicit comparison. In this context, digital systems can provide examples or reformulations that encourage learners to recognise patterns. Research suggests that such variation, when embedded in reflective tasks, can contribute to grammatical awareness (Lee, 2021; Godwin-Jones, 2018).

Semantics – Meaning and nuance. Semantics concerns the meaning of words and sentences. Related languages present particular semantic pitfalls: words may resemble one another while carrying different meanings. These are the classic *false friends*, which can lead to persistent misunderstandings.

Dutch	English	Spanish	Italian	Portuguese	Comment
voelen	to feel	*sentir*	*sentire*	*sentir*	Spanish: also 'to smell'
ouder	parent	*padre / madre*	*genitore*	*pai / mãe*	Italian *parente* = relative
licht	light	*luz*	*luce*	*luz*	All three: also figuratively 'insight'
leuk	nice / fun	*divertido*	*divertente*	*divertido*	Italian: more limited use for people

Semantic awareness develops when learners actively learn to recognise and articulate differences in meaning and usage. Supporting digital tools can provide paraphrases or context-sensitive alternatives that facilitate this process (Reynolds, 2023; Zhai, 2022).

Pragmatics – Language use in context. Pragmatics concerns appropriateness and socially suitable language use. Language is not only correct or incorrect, but also appropriate or inappropriate in a particular situation. This domain is crucial to genuine communicative competence, but is often neglected. Example situation: ordering something.

	Dutch	English	Spanish	Italian	Portuguese
Direct request	*Ik wil een koffie.*	I want a coffee.	*Quiero un café.*	*Voglio un caffè.*	*Quero um café.*
Polite form	*Mag ik alstublieft een koffie?*	May I have a coffee, please?	*Me gustaría un café.*	*Vorrei un caffè.*	*Gostaria de um café.*

Variation is essential for this kind of context-sensitive language use. Learners may benefit from digital tools that present alternatives for formal and informal usage. Reflective discussion of such variants can contribute to pragmatic awareness (Taguchi, 2015; Lee & Drajati, 2019).

Concluding remark. What is commonly understood as 'grammar' — sentence structure, inflection and word order — falls largely within syntax and morphology in this classification. This dissertation deliberately adopts the more fine-grained division into five domains because it helps language learners organise similarities and differences more effectively. Structuring language input along these lines provides a manageable framework that encourages comparison without requiring rules to be explained explicitly. The domains thus function as cognitive labels for errors, discoveries and patterns, and thereby form the cognitive foundation of the polyglot learning programme developed in the remainder of this dissertation.

§2.5 Interlingual Comparison and Transfer: Synergy or Interference?

Introduction: multilingualism as both mirror and pitfall. Anyone learning several languages simultaneously is constantly confronted with differences between them. Every new word, grammatical form and sentence structure is automatically compared with previously acquired languages — consciously or unconsciously. This process of comparison contains both an opportunity — structural recognition, reinforcement of patterns and semantic deepening — and a risk — overgeneralisation, interference and confusion between systems.

In the literature, this mechanism is referred to as *transfer*: the transfer of knowledge or skills from one language to another. Transfer is an inevitable cognitive phenomenon in multilingual language acquisition (Odlin, 1989). The central assumption in this dissertation is that transfer — provided that it is consciously guided — is not merely a side effect, but an instructional principle. Two main forms of transfer can be distinguished: 1) *Positive transfer:* knowledge from language A helps in correctly interpreting or producing language B (for example, recognising the shared stem structure in *información – informazione – informação*). 2) *Negative transfer or interference:* knowledge from language A is incorrectly applied in language B (for example, placing a pronoun in the wrong position in a Spanish sentence by analogy with Dutch). Both forms of transfer are inevitable in simultaneous language acquisition. The learning programme developed in this research therefore does not seek to avoid transfer, but to activate it: similarities and differences are deliberately made visible in order to structure the learning process.

Mechanisms of transfer: what is transferred, and how? Transfer can occur at several levels (Ringbom, 2007; Jarvis & Pavlenko, 2008): 1) Lexical (word level): words resemble one another (*información, informazione, informação*); 2) Morphological: verb endings or gender rules are assumed to be universal; 3) Syntactic: word order or the placement of clitics is incorrectly transferred; 4) Semantic: false friends and differences in meaning cause confusion (*divertido, divertente*); 5) Pragmatic: politeness strategies or direct forms of expression are transferred. In practice, these levels of transfer are often intertwined: form and meaning influence one another, and transfer rarely manifests itself in pure categories (Jarvis & Pavlenko, 2008; Odlin, 2003). Crucially, language learners generally begin to reflect on transfer only when something goes wrong — for example, in response to errors, confusion or contradictory input. From this perspective, errors are not merely obstacles, but also instructional windows: opportunities to reconsider or clarify underlying structures.

Example: correct transfer versus interference

Meaning	Dutch	English	Spanish	Italian	Portuguese
I see him tomorrow	*Ik zie hem morgen*	I see him tomorrow	*Lo veo mañana*	*Lo vedo domani*	*Eu o vejo amanhã*
Interference error	*Ik hem zie morgen*	*I him see tomorrow*	*Veo lo mañana*	*Vedo lo domani*	*Vejo o amanhã*

These examples make clear that the same word order cannot be transferred universally. Correct use of clitics — pronominal objects — requires language-specific placement, which, precisely because of the similarities between the languages, can easily be transferred incorrectly.

Instructional implication: transfer as a learning resource. In this research, transfer is not treated as a problem to be avoided, but as an indication of where and how languages interact — and therefore as a point of departure for reflection and guidance within the learning process. When a language learner makes an error resulting from the incorrect transfer of a familiar structure, this provides an ideal moment for explicit reflection: Why does this not work in Portuguese? What changes if I apply the Italian pattern to Spanish? Which construction is possible in all three languages, and where is it not?

AI applications such as GPT can play a supporting role in this learning process. Because the model can provide alternative formulations of the same meaning in different languages on request, structural similarities and differences become visible. Research indicates that this form of language modelling — provided that it is appropriately guided — can support the development of cross-linguistic awareness (Liu et al., 2023; Reynolds, 2023). By encouraging language learners to compare actively, formulate their own assumptions about linguistic structures, and ask questions of the model through prompts, transfer is not suppressed but instead transformed into a source of reflection and insight. The underlying assumption is that multilingual acquisition proceeds more rapidly and at greater depth when the learning process is structured around interlingual comparison: the systematic juxtaposition of meaning equivalents across several languages, including their grammatical form, semantic nuance and pragmatic implications.

Interlingual strategies as a concretisation of transfer mechanisms. This subsection describes the system of interlingual strategies developed within this research as a concrete operationalisation of the transfer mechanisms discussed in this section. The strategies are designed to activate conscious positive transfer, reduce negative transfer or interference, and systematically stimulate metalinguistic awareness. They therefore provide a structured instructional and analytical framework through which interlingual comparison, as a driving mechanism of polyglot language acquisition, is operationalised both in the learning design (Chapter 3) and in the evaluation strategy (Chapter 4).

The system of eight interlingual strategies (M, F, G, S, T, D, W, Z) was developed in this research as a structured concretisation of transfer mechanisms described in the literature. The eight strategies constitute an instructional-analytical framework derived from existing theoretical concepts concerning transfer, metalinguistic awareness and contrastive analysis (Odlin, 1989; Ringbom, 2007; Jessner, 2006; Cenoz & Gorter, 2011), but have not previously been positioned as a fixed system in this form.

The strategies can be summarised as follows:

- **M (Morphological):** awareness of similarities and differences in word formation — stems and affixes — aimed at the inductive recognition of morphological patterns.
- **F (Phonetic):** identification of similarities and differences in sound patterns to support listening comprehension and accurate pronunciation.
- **G (Graphic):** recognition of visual similarities in written form and spelling, supporting word recognition and orthographic accuracy.
- **S (Systematic suffix):** explicit analysis of systematic correspondences in word endings — for example, Spanish *-ción*, Italian *-zione*, Portuguese *-ção* — to support semantic and syntactic prediction.
- **T (Transfer):** conscious application of familiar patterns from a related language to a new language, accompanied by reflection on correctness and contextual appropriateness.
- **D (Divergence):** awareness of structural or lexical differences that may cause interference; a strategy for avoiding false friends and structural pitfalls.
- **W (Alternating consonants):** recognition of systematic sound or letter alternations in related words — for example, Spanish *b*, Italian *v*, Portuguese *v* — to support active vocabulary development.
- **Z (Sentence structure):** analysis of syntactic similarities and differences to support accurate sentence formation and syntactic transfer.

The relatedness between the Romance languages makes these strategies not only applicable, but particularly instructive. They are rooted in a shared grammatical and lexical inheritance extending back to Latin. This historical point of departure can provide a **conceptual key** for understanding how structural similarities between languages increase cognitive accessibility. By recognising related forms and interpreting their divergences, the learner comes to understand relationships between languages not through rules, but through analogies. In this sense, Latin has instructional potential as a **mental reference framework**: it provides an organising perspective on formal relatedness, semantic stability and systematic variation. This makes clear how relatedness itself can function as a learning resource — an idea that is also relevant to the theoretical discussion of generative models in the following section.

In this dissertation, these strategies function as a conceptual framework that is embedded in tasks and reflection formats within the learning design (Chapter 3), and serves within the analysis (Chapter 4) as a basis for interpreting output, reflections and error patterns. Empirically, the research tests the extent to which explicit use of these strategies contributes to faster acquisition (subquestions 1 and 3), more accurate use of interlingual patterns (subquestions 2 and 3), and increased metalinguistic awareness (subquestion 2).

§2.6 Metalinguistic Awareness as a Catalyst

Introduction: from language use to thinking about language. Whereas much language education focuses on learning to use language correctly, this research goes one step further: it assumes that language acquisition can be accelerated and deepened when the language learner also learns to think about language as a system. In the linguistic and psycholinguistic literature, this capacity is referred to as *metalinguistic awareness*: the ability to analyse, structure, identify and compare language independently of its immediate communicative function (Gombert, 1992; Bialystok, 2001).

Metalinguistic awareness involves more than knowing grammar. It concerns the ability to ask questions such as: Why does the verb appear at the end? How does the meaning change if I choose a different verb form? Is this polite in Italian, or too direct? In the context of polyglot language acquisition, this ability is crucial. Simultaneously learning three related languages requires the language learner to learn to distinguish between systems, recognise changes in meaning resulting from changes in form, and develop strategies for dealing with uncertainty and errors.

Metalinguistic awareness in the literature. Research indicates that multilingual speakers generally perform better than monolingual speakers on metalinguistic tasks (Jessner, 2006; Galambos & Goldin-Meadow, 1990). This is not coincidental, but a systemic consequence: learning several languages requires users to compare structures, differentiate between contexts and classify linguistic forms. Jessner (2008) argues that metalinguistic reflection is not only a consequence of multilingualism, but also a condition for acquiring languages efficiently. This corresponds to the assumption underlying this research: polyglottism requires active thinking about language, not merely language use.

The role of LLM applications in metalinguistic reflection. This research assumes that AI applications such as GPT can support metalinguistic reflection by enabling language learners to compare sentence structures, reformulate utterances and make differences in meaning explicit. This assumption is consistent with earlier research on AI-facilitated language feedback and pattern recognition (Lee, 2021).

Examples of the role an LLM can play include enabling the language learner to *compare sentence structures across languages*; providing immediate feedback on errors, including paraphrases and corrections; and generating alternatives that encourage reflection on tone, tense, order and context. Examples include: “What is the difference between *conocer* and *saber*?”, “Is *vorrei* more polite than *voglio*? And why?” and “How do I place the pronoun correctly in Portuguese?”

Rather than merely providing explanations, an LLM can create opportunities for the language learner to discover, question and compare independently. This form of interaction — in which language functions both as input and as an object of reflection — can stimulate metalinguistic thinking in practice.

Instructional implication. In the learning programme designed in this research, metalinguistic awareness is not treated as a by-product, but as an explicit learning objective. This is achieved not by teaching terminology, but by creating situations in which language prompts reflection: through interlingual comparison, error analysis, and LLM-generated counterexamples and nuanced variants. The assumption is that language acquisition is strengthened when learners

are actively supported in formulating questions about language, recognising structural variation and explaining differences in meaning. This assumption is consistent with constructivist approaches in which learning develops through reflection on meaningful input (Tomasello, 2009; VanPatten, 2015). In this context, the LLM functions as an interactive tool that can support this reflective process, for example by generating examples, proposing reformulations or exposing contrasts. The model is not used as an authoritative source of knowledge, but as a means of stimulating awareness and pattern recognition.

GPT within the Zone of Proximal Development (ZPD). In terms of Vygotsky's theory of the Zone of Proximal Development (ZPD; 1978), GPT can function as a form of scaffolding — temporary support pitched just above the language learner's current level. The model can provide suggestions, alternatives and reformulations that the learner would not yet have been able to produce independently, but can understand and process with some guidance. By engaging actively with this output — for example, by using targeted prompts or examining differences between formulations — the learner operates within precisely the learning space in which development can occur: not so difficult as to cause blockage, but not so easy as to cause boredom.

The central point of this dissertation is therefore not that GPT is a language model, but that, as an instructional medium, it can provide a platform for active learning in which learners make their own choices *(self-direction),* infer rules and patterns from examples themselves *(induction),* and consciously reflect on how language is structured *(reflection).* These processes can only operate effectively when the learning programme is deliberately designed around clear objectives, boundaries and attention to transfer between languages.

§2.7 The LLM as an Instructional Instrument: Possibilities and Limitations

Generating language is not the same as understanding language. GPT models, as members of the family of large language models (LLMs), produce text on the basis of statistical probabilities (Brown et al., 2020; Bender & Koller, 2020). They do not possess semantic understanding, the ability to assess truth, or communicative intention, but reconstruct language as a probabilistic system. It is precisely this limitation that makes them theoretically interesting: they reproduce the structural dimension of language, not its meaning-bearing dimension. In doing so, they shed new light on the question of what learning entails when language acquisition is understood as the recognition of patterns rather than the application of rules.

Within theoretical frameworks of emergent and usage-based learning (Ellis, 2003; Tomasello, 2009), repeated exposure to variation is regarded as a driving force of acquisition.

From this perspective, LLMs can be regarded as external sources of variation that offer a virtually unlimited reservoir of possible combinations. They do not function as knowing subjects, but as representations of a linguistic environment in which patterns are statistically reconstructed. In this capacity, they can perform several functions relevant to theorising about language acquisition. A model can, for example, function as a pattern generator by producing grammatically correct constructions in numerous variants. The learner is thereby exposed to subtle shifts in form and meaning that may facilitate emergent learning. The same model can

also function as a reformulator, capable of rewriting existing sentences in different registers or styles, thereby making pragmatic nuance and collocation visible. It can likewise function as a comparator by generating parallel sentences in different languages and thus exposing structural relationships that are crucial to understanding transfer and interference (Jessner, 2006). Finally, in the spirit of Vygotsky's (1978) Zone of Proximal Development, an LLM can provide temporary scaffolding by completing incomplete constructions or reformulating incorrect sentences. Such interventions may mark the transition between what the learner can accomplish independently and what becomes possible with minimal assistance, provided that these comparisons are accompanied by reflection (Flavell, 1979; Schraw & Dennison, 1994).

The usefulness of LLMs as an instructional concept is, however, constrained by the same statistical principles that make them powerful. In infrequent structures, inconsistencies may occur (Liu et al., 2023), and in languages with limited representation in the training data, such as Swahili or Yoruba, morphological patterns are often modelled incompletely. For strongly represented languages, including Spanish, Italian and Portuguese, such shortcomings are less pronounced, but pragmatic sensitivity remains problematic: registers, forms of politeness and irony are only partially distinguished. Moreover, the abundance of available variants can lead to cognitive overload, making it difficult for the learner to isolate the relevant pattern. The same generative abundance that can stimulate learning may therefore also obscure understanding.

In summary, within the theoretical framework of language acquisition, LLMs function as experimental models that simulate exposure, variation and temporary support without intention or understanding. They make visible how frequency and scaffolding may contribute to structural understanding, while also revealing the limits of non-intentional language production and pragmatic interpretation.

§2.8 Synthesis: Barriers and Accelerators in AI-Facilitated Polyglot Language Acquisition

Introduction: overview of the tensions. The preceding sections have developed a theoretical framework that positions the use of AI applications such as LLMs in the simultaneous learning of three related languages — Spanish, Italian and Portuguese — in relation to models of language acquisition, cognitive mechanisms and principles of instructional design. This concluding section summarises the principal accelerating factors and barriers, providing a bridge to the learning design presented in Chapter 3.

Accelerators:

1. **Tailored, pattern-rich input.** An LLM can provide rich, repeatable and contextually adaptable example sentences. This supports usage-based learning, in which the language learner infers structural patterns without explicit instruction (Ellis, 2003).
2. **Comparative output across languages.** Generating the same meaning in several languages can reveal systematic similarities and differences — the basis for interlingual transfer (Ringbom, 2007).
3. **Immediate, non-judgemental feedback.** An LLM can provide error analysis and reformulation without social pressure, making self-correction more accessible (Lynch, 2001).

4. **Stimulation of metalinguistic awareness.** Through examples, paraphrases and reformulations, an LLM can strengthen the ability to reflect on language, particularly in domains such as morphology, syntax and pragmatics (Jessner, 2006).
5. **Scaffolding within the Zone of Proximal Development (ZPD).** An LLM can function as temporary support for reflection, hypothesis formation and self-direction (Wood, Bruner & Ross, 1976; Vygotsky, 1978).

Barriers:

1. **Inconsistent or incorrect output.** An LLM can generate grammatically or pragmatically incorrect forms, particularly in less dominant languages or when dealing with more complex structures.
2. **Inability to provide explanation or feedback at a meta-level.** The model does not provide principled explanations or analyses; without external guidance, interpretation is lacking.
3. **Cognitive overload during simultaneous processing.** The abundance of alternatives and translation suggestions can make it difficult to maintain an overview, particularly for beginners learning three languages simultaneously.
4. **The black-box problem.** Because the operation of the model is not transparent, error patterns may persist without being recognised (Bender & Koller, 2020).
5. **Risk of passive dependence.** Without targeted tasks, an LLM becomes a provider of answers rather than a stimulus for insight. The model does not independently direct learning objectives or provide adaptive feedback (Lim et al., 2023), making guidance necessary to address misconceptions.

Core insight: design is essential. This overview leads to a fundamental conclusion: an LLM is not an instructional system that independently formulates learning objectives or corrects errors, but a generative medium that can contribute to language acquisition under specific conditions. Whether it functions as an accelerator or a barrier is not inherent in the model itself, but depends on the instructional design within which it is embedded. An LLM is most effective in learning environments in which input is purposefully linked to specific linguistic domains; tasks encourage comparison and self-correction rather than mere imitation; errors are used as opportunities for reflection; and the learning objective is not perfect mastery, but conscious and flexible multilingualism.

§2.9 Language Aptitude, Linguistic Complexity and the Rationale for Selecting Three Related Romance Languages

Whereas Chapter 2 has so far described the mechanisms that can accelerate or impede polyglot language acquisition, this concluding section focuses on the cognitive conditions on the part of the language learner and on the structural properties of the languages themselves, which together determine whether a simultaneous learning programme is actually feasible.

Capacity for language acquisition. Research on language acquisition shows that the feasibility of a learning task is partly determined by the extent to which the learner's cognitive capacities correspond to the structural demands imposed by the target language. Within the classical aptitude tradition (Carroll & Sapon, 1959; Carroll, 1981; Skehan, 1998, 2016; Meara

& Rogers, 2019), language aptitude is not understood as a single ability, but as a configuration of interrelated components: the capacity to encode phonological patterns reliably, sensitivity to structural configurations ('grammatical sensitivity'), the ability to retain new vocabulary ('rote memory'), and the ability to form autonomous generalisations on the basis of variation in input (inductive learning). More recent studies (Li, 2015) show that these components are taxed to different degrees depending on the structural properties of the target language, and that the interaction between aptitude and language characteristics is crucial to the speed with which patterns are acquired.

These cognitive components correspond to structural domains in which languages may vary. Phonological coding is more heavily taxed in languages in which spelling and pronunciation are relatively unpredictable, because the learner cannot anticipate sound structure from the written form; 'grammatical sensitivity' is more critical in languages with rich inflectional morphology or less rigid word order; 'rote memory' is taxed more heavily when the lexicon is highly non-transparent; and inductive ability becomes essential in languages displaying substantial context-dependent variation or in which relevant patterns are only implicitly available. The relative difficulty of a language — and therefore the demands placed on these aptitude components — has been shown to correlate with structural distance between languages (Kusters, 2003; Bentz & Berdicevskis, 2016), a distance that is quantified in recent computational models by means including phonotactic predictability, morphological prediction and measures of lexical distance (Bentz et al., 2017; Serva & Petroni, 2008; Mayer & Cysouw, 2012).

Linguistic complexity. In computational linguistics, structural distance is generally modelled as a function of several underlying variables that together determine linguistic complexity (Bentz & Berdicevskis, 2016; Kusters, 2003). Four dimensions are particularly relevant to this research.

The first is *phonological predictability (P):* the extent to which the sound structures of a language are systematic and the relationship between spelling and pronunciation is consistent. This predictability is often quantified through *phonotactic entropy*, with languages characterised by stable sound patterns and transparent spelling–sound correspondences displaying lower entropy values (Bentz et al., 2017).

The second dimension is *morphological transparency (M):* the regularity with which word forms and inflections are constructed. Computational models describe this in terms of morphological prediction: the extent to which word forms can reliably be derived from stems and affixes. Typological parameters from the World Atlas of Language Structures (WALS) indicate that languages with relatively few irregular word forms — and limited allomorphic morphology — exhibit greater predictability and lower cognitive demands (Bickel & Nichols, 2013).

The third dimension is *syntactic stability (S):* the consistency of word order and dependency relations within sentences. Languages with a stable SVO pattern and limited variation in clitic placement display greater syntactic predictability; this is typologically supported by parameters relating to dependency direction and basic word order (Dryer & Haspelmath, 2013; Futrell et al., 2015).

The fourth dimension is *lexical relatedness (L):* the extent to which word forms correspond across languages, either historically or through systematic form–meaning relationships. This

relatedness is generally measured using Levenshtein distances, cognate analyses and cross-lingual vector representations (Serva & Petroni, 2008; Ringbom, 2007). A high degree of lexical similarity makes it possible to infer new words from existing word families, substantially reducing demands on memory processes.

In combination, P, M, S and L determine the effective structural distance between languages. Across these four dimensions, the three Romance languages display mutually predictable phonotactics, substantial morphological parallels, a stable basic syntactic structure and a high proportion of transparent vocabulary (Bonvino, 2009). This configuration explains why the simultaneous acquisition of Spanish, Italian and Portuguese is cognitively less demanding than the acquisition of three typologically disparate languages: overall structural distance is low across all four dimensions, while the remaining divergent patterns are sufficiently systematic to support inductive learning.

Linguistic complexity in polyglot learning. Within this structural proximity, an important interaction arises with the aptitude components described above. The phonological burden is reduced because the three languages use closely related sound systems. The lexical burden is lower because most word families have transparent or regular correspondences, thereby placing fewer demands on 'rote memory'. Grammatical sensitivity does not continually have to be directed towards three independent paradigms, but operates within a shared grammatical configuration characterised by parallel conjugation patterns and a largely congruent basic syntax. In this context, inductive ability is not impeded by noise or irregularity, but is instead engaged by the possibility of abstracting systematic correspondences and divergent elements. Simultaneous acquisition of these three languages therefore does not require a threefold cognitive investment, but rather a shared deployment of the same underlying capacities.

Rationale for the selection of the target languages. The Spanish–Italian–Portuguese combination was selected because the three languages display a low structural distance in absolute terms and therefore create a manageable cognitive burden within a simultaneous learning programme. Their shared phonological, morphological and syntactic basic structures increase pattern predictability, while the remaining differences are small but sufficiently systematic to make comparison meaningful. Their mutual lexical relatedness is also exceptionally high, substantially facilitating vocabulary acquisition and consolidation. Finally, these three languages are generated relatively consistently by contemporary language models, which is relevant to this research because example sentences and comparative material can be produced reliably in all three languages. Taken together, these characteristics make this triad a typologically and cognitively well-founded case for examining the simultaneous acquisition of related languages.

Table 2.9 shows that Spanish, Italian and Portuguese display a relatively low structural distance. The values in this table are derived from independent typological and computational descriptions of each of the three languages, as documented in WALS, studies of phonotactic entropy (Bentz et al., 2017) and measurements of lexical distance (Serva & Petroni, 2008). The classification does not compare the languages against a single reference point, but presents their absolute characteristics across four structural dimensions: phonology, morphology, syntax and cognate structure. These dimensions were selected because they correspond directly to the MLAT components that determine the cognitive demands of simultaneous language acquisition (Carroll, 1981; Skehan, 2016). Across all four dimensions, Spanish, Italian and Portuguese exhibit relatively high levels of predictability and transparency compared with other typological

clusters. In absolute terms, they therefore constitute a suitable triad for simultaneous polyglot acquisition within an accelerated programme.

Dimension	Definition	Spanish	Italian	Portuguese	Instructional implication for simultaneous acquisition
P – Phonological predictability	The extent to which spelling and pronunciation correspond predictably; stability of sound combinations	High – highly predictable spelling–pronunciation relationship, simple sound combinations	High – predictable sound patterns and limited pronunciation variation	Medium – clear basic patterns, but greater pronunciation variation due to vowel reduction	A predictable relationship between spelling and pronunciation reduces the cognitive burden of word recognition, leaving greater capacity for comparing patterns across the languages.
M – Morphological transparency	Regularity and predictability of inflectional and conjugational patterns	High – largely regular verb conjugation and morphology	High/Medium – richer morphology but highly predictable patterns	Medium – predominantly regular morphology with some deviations	Regular morphological patterns are recognised more quickly, making differences between the languages more likely to be perceived as systematic rather than as isolated exceptions.
S – Syntactic stability	Consistency of word order and grammatical relations	High – stable SVO word order and predictable placement of short forms such as pronouns	High – comparable basic structure and little unexpected variation	High/Medium – stable basic structure with greater variation in pronoun placement	Consistent sentence structures reduce learning-irrelevant noise, making comparisons between sentence structures in the three languages less demanding.

Dimension	Definition	Spanish	Italian	Portuguese	Instructional implication for simultaneous acquisition
L – Lexical relatedness	The extent to which word forms in the three languages correspond systematically in pairwise comparison (Sp–It, Sp–Pt, It–Pt), based on regular sound and form correspondences	Very high – many shared word forms within the Romance family	Very high – strong correspondence with Spanish and a broad cognate base	High – largely transparent word forms with predictable shifts	Shared word forms reduce memory demands and make it possible to infer new words from existing knowledge, supporting simultaneous vocabulary development.

Table 2.9: Characteristics and relationships of the Romance languages.

This table presents four absolute dimensions of complexity, only one of which — lexical relatedness — directly indicates the mutual relatedness of the three languages. High predictability on the other three dimensions — phonological, morphological and syntactic — primarily facilitates the inductive learning of each language individually. What the table does not show, but what is well documented in the comparative literature on the Romance languages (Bonvino, 2009; Maiden, 2016; Posner, 1996), is that Spanish, Italian and Portuguese also share *substantial structural similarities* in their sound inventories (phonology), inflectional patterns (morphology) and basic word order (syntax). This combination of demonstrable lexical relatedness between the three languages, high individual predictability across key dimensions of complexity, and shared phonological, morphological and syntactic structures makes this Romance triad particularly suitable for inductive, simultaneous polyglot acquisition.

This combination of absolute predictability, documented structural similarities and manageable cognitive demands provides the theoretical point of departure for Chapter 3, in which these insights are translated into a concrete and testable learning design.

Chapter 3 – Learning Design

§3.1 Introduction – From Theory to Design

In the preceding chapter, the theoretical plausibility of language acquisition without explicit grammar instruction was established. In this chapter, the perspective shifts from explanation to design: how can the principles of emergence, frequency sensitivity and interlingual transfer described above be translated into a feasible learning programme that enables the simultaneous acquisition of three Romance languages? The purpose here is not to repeat the theory, but to make explicit the design decisions that justify the transition from theoretical possibility to instructional reality.

The learning design constitutes the empirical core of this research. It functions as an intermediate layer between the theoretical model and practical implementation: a formal structure in which hypotheses concerning polyglot language acquisition are translated into concrete learning conditions, materials and instruments. The design must therefore be not only effective but also explicable — every design choice must be traceable to a theoretical assumption, while every assumption must, conversely, be given a testable translation within the learning design.

Within the broader framework of this research, three functions are fulfilled here: 1) an **epistemological function** — concretising the core hypothesis that emergent pattern learning, guided by frequency and variation, is sufficient to achieve A2 level in three related languages; 2) an **instructional function** — specifying the building blocks of the learning programme: scaffolding, the micro-curriculum, design rules for learning materials and the role of AI-facilitated support; and 3) a **legitimating function** — providing the rationale for why this particular combination of design choices is theoretically consistent and empirically defensible.

By placing theory, design and testability on a single continuum, this chapter marks a **key transition within the dissertation**: it becomes possible to see how an abstract model of language acquisition can be translated into a reproducible and verifiable learning design. Whereas Chapter 2 provided the cognitive rationale, Chapter 3 provides the instructional rationale — thereby establishing the foundation on which the learning book and empirical analysis are built.

Although the primary purpose of this chapter is to provide the theoretical justification for the learning design, §3.8 marks a first shift towards the empirical dimension: not through the analysis of learning outcomes, but through systematic reflection on the difficulties encountered during the design and production phase itself. These design experiences lead to refinements of the original design and thereby constitute an empirical layer *avant la lettre*.

§3.2 Design Philosophy – Learning Like an LLM

The design philosophy underlying this learning programme starts from a fundamental shift in thinking about language acquisition: learning is no longer understood as the internalisation of rules, but as the statistical recognition of patterns in meaningful input. This approach is consistent both with usage-based learning (Bybee, 2010; Ellis, 2003) and with insights from computational linguistics and neural language models (Jurafsky & Martin, 2024; Goldberg, 2019). Like an LLM, the language learner abstracts structural relationships from frequent examples without these ever being presented explicitly as rules.

The grammar-free paradox. Explicit grammar instruction is deliberately not taken as the point of departure for the learning design. The intention is for structural understanding to emerge through repeated exposure to semantically rich, parallel sentences. This choice appears paradoxical: how can patterns be learned without knowing the rules? The answer lies in the distinction between rule explanation and rule sensitivity (Ellis & Larsen-Freeman, 2006). Whereas the former requires explicit instruction, the latter relies on distribution-sensitive perception: the ability to register which forms occur in which contexts. This capacity is deeply rooted in human cognition and, as discussed in Chapter 2, forms the basis of emergent language learning (Tomasello, 2008; Kuhl, 2004).

Frequency, variation and emergence. Within the design, this sensitivity to emergent learning is activated through a controlled balance between frequency and variation. Each new lexical unit appears in multiple contexts across three related languages, requiring the learner to form generalising abstractions. Research on construction grammar indicates that meaningful repetition across varied contexts accelerates the formation of cognitive schemas (Bybee, 2010; Goldberg, 2019). The same principles are applied here to simultaneous multilingualism: the three languages function as three observation points on a single underlying pattern.

The design makes use of *semantically equivalent triplets*: for each lemma, three parallel sentences — Spanish, Italian and Portuguese — with the same semantic core but different forms. This form of controlled variation creates an internal tension between similarity and difference — precisely the cognitive context in which pattern recognition is assumed to yield the greatest return (De Bot & Lowie, 2020).

Learning as a generative model. The comparison with LLM-like systems is not merely metaphorical, but functions as an instructional model. Both humans and language models learn through *next-token prediction*: anticipating the most probable next form within a given context. The difference is that the human learner, unlike the model, possesses semantic awareness and the capacity to recognise intention (Hauser, Chomsky & Fitch, 2002). The learner can therefore not only predict but also *understand*. The learning design exploits this difference: learners are exposed to LLM-like distributions, but anchor the resulting patterns through reflective tasks that situate meaning and use. The result is a hybrid learning process: emergent learning supported by conscious comparison and reflection.

Rationale for the design philosophy. The choice of a grammar-free, distribution-driven design is scientifically justified by converging evidence from three domains: 1) **Cognitive linguistics:** emergent pattern learning provides an explanation for the acquisition of syntactic regularities without explicit instruction (Ellis, 2003; Tomasello, 2008); 2) **Usage-based models:** frequency and type variation determine the rate at which constructions become entrenched (Bybee, 2010; Goldberg, 2019); and 3) **Computational linguistics:** neural language models provide empirical

evidence that statistical exposure can be sufficient for grammatical structure to emerge (Jurafsky & Martin, 2024; Tenney et al., 2019).

The learning design developed in this dissertation constitutes the human counterpart of that principle: a controlled environment in which distributional exposure, rather than the explicit inculcation of rules, functions as the driving mechanism of acquisition. The design philosophy thereby provides the justification for the transition from theoretical plausibility (Chapter 2) to instructional feasibility (Chapter 3).

§3.3 Linguistic Foundation – Latin as an Interlingual Key

The decision to teach Spanish, Italian and Portuguese simultaneously is based on more than practical considerations. It is rooted in a deeper linguistic coherence: despite their divergent development, the three languages remain variants within a single structural field. Vulgar Latin, from which they descend, is not a learning objective within this design, but functions as an explanatory and organising principle. Whereas traditional classical-language education approaches Latin as a grammatical model — a system intended to provide insight into linguistic structure through rules and paradigms — this research uses it as an *interlingual key*: a conceptual reference through which the learner can understand structural relatedness rather than merely experience it.

Latin is not present in this design as a language to be mastered, but as an underlying pattern that explains the recognisability of relationships between the three modern languages. Although centuries of phonological erosion, morphological reduction and syntactic reconfiguration have caused the Romance languages to diverge, they still share a network of correspondences that appeals to linguistic intuition. The learner recognises that corresponding sounds, stems and endings recur in different forms: *scrib-* resonates in *escribir*, *scrivere* and *escrever*; the Latin suffix *-tio* survives in *-ción*, *-zione* and *-ção*. Such patterns do not need to be explained in order to be effective: they are recognised rather than learned. Genealogical relatedness thus provides the background against which pattern recognition becomes possible, without requiring the learner to engage with the complex grammar of the ancestral language of the Romance family.

From this perspective, Latin provides both a cognitive and a scientific key. Cognitively, it offers a form of implicit orientation: learners are not learning three arbitrary languages, but three variants of the same structural logic. By experiencing this underlying systematicity, they can draw analogies and identify differences more precisely without explicit knowledge of rules. Scientifically, the concept of a common origin legitimises the choice of simultaneous acquisition. It explains why comparative learning is not arbitrary, but rests on structural continuity that has been extensively documented in linguistic research (Maiden, 2011; Bonami & Boyé, 2022). The learning design thereby becomes more than an instructional experiment: it constitutes a controlled re-enactment of the historical divergence of Latin itself — a process through which the learner proceeds in reverse.

The interlingual strategies used in the learning book are the operational manifestation of this key. They transform genealogical relatedness into observable learning moments. When learners recognise similar word stems or sound correspondences, they are effectively reconstructing a shared linguistic network, albeit implicitly. The strategies are therefore not an attempt to teach Latin, but a means of making the residue of Latin form-patterning visible again. They articulate

what the brain already does: registering similarities, systematising them and using them for prediction and transfer.

This approach is consistent with research on *cognate facilitation* and *cross-linguistic transfer*, which convincingly demonstrates that learners of related languages benefit from shared morphological and semantic networks (de Groot & Keijzer, 2000; Otwinowska, 2016; Bono & Varela, 2021). By systematically organising this natural relatedness, transfer is not left to chance but incorporated into the design. The three languages function as mirrors of a single system: difference becomes instructive because it presupposes similarity.

Within this framework, Latin fulfils neither a normative nor an instructional role, but provides a metalanguage in which the relationships between the three languages can be expressed conceptually. It functions as a bridge between historical language evolution and contemporary pedagogy: a silent presence that explains why patterns can be recognised without prescribing how they should be learned. The genealogical framework is thus not elevated to the status of learning material, but serves as an epistemological justification for the design itself. The learner does not learn *from* Latin, but *through* a structure whose shadow still falls across the modern Romance languages — and it is precisely this shadow that makes simultaneous learning plausible.

§3.4 Instructional Architecture of the Learning Design

The learning design presented in this dissertation is not composed of separate instructional principles, but of a coherent configuration of mutually dependent processes. Together, these form a self-reinforcing system in which pattern-rich exposure, emergent learning, reflection and iterative repetition feed into one another cyclically. The architecture is designed to show how the mechanisms described in the theoretical framework (Chapter 2) can be translated into a practically feasible learning environment that is both testable and reproducible.

The point of departure is simultaneous exposure to three related languages. This concurrent exposure generates rich and variable input in which similarities and differences become visible at the same time. The learner is not presented with three isolated language codes, but with a continuum of variation within which patterns can be detected automatically. This environment provides the foundation for all the other design principles. Without parallel input, there would be no basis for comparison, and without comparison, no awareness of structural similarity or divergence.

Within this rich input environment, grammar-free acquisition can occur: structural knowledge emerges not from explanations of rules, but from the recognition of distributional patterns. The sentences in Spanish, Italian and Portuguese function as a dynamic corpus in which frequency and context organise repetition. This form of emergent learning would, however, remain unguided without AI-based control to keep variation within defined boundaries. The use of LLMs in material production ensures that each triplet is semantically equivalent, morphologically parallel and syntactically natural. This creates the instructional conditions for emergence: sufficient variation to elicit generalisation, but sufficient order to prevent confusion.

Temporally, the entire learning environment is structured through *iterative scaffolding*. The twelve weeks of the programme follow a logical progression from perception to production:

what is recognised phonologically and lexically in the first phase later reappears in more complex syntactic and pragmatic contexts. Each cycle reinforces the previous one, so that patterns are not built up linearly but deepened in a spiral. This temporal structure is essential: without repeated and varied encounters with the same structures, emergent knowledge would quickly dissipate.

When these principles operate together, a *didactic ecosystem* emerges that regulates itself. Simultaneous exposure provides the raw material for pattern formation; grammar-free acquisition translates that input into implicit structure; AI-based control maintains order within the variation; reflective anchoring consolidates the acquired knowledge; and iterative scaffolding renews the cycle in increasingly rich contexts. Each principle therefore functions as a precondition for the next. Remove one, and the system loses its stability: without AI consistency, the structure becomes blurred; without reflection, learning remains superficial; without scaffolding, temporal coherence disappears.

The strength of this architecture lies precisely in its internal dynamics. At an instructional level, the design reproduces what is described in neural and computational models of language acquisition: a network of recurrent processes in which output continuously becomes input for a subsequent phase. Learning is not a linear progression here, but a closed cycle of exposure, recognition, consolidation and expansion. The design thereby provides a concrete illustration of how the theoretical assumptions concerning emergence, distribution and frequency discussed in Chapter 2 can be translated into an empirically testable instructional system.

In other words, the whole is more than the sum of its parts. Together, the five design principles form an adaptive mechanism that allows grammatical knowledge to emerge where no explicit instruction is provided, while maintaining order where variation is the engine of learning. This architectural coherence makes the learning design not only feasible, but also explicable: each learning activity contributes to a larger, self-reinforcing pattern. The scientific core of this research lies in that self-regulating structure.

§3.5 Design Rules for Practice Materials and Variation

The learning design rests not only on abstract principles, but also on a fine-grained system of rules governing the form, content and distribution of the practice materials. The quality of the design depends on the extent to which these rules succeed in combining the necessary variation with structural consistency. Variation is not an incidental by-product of the learning process, but an instructional strategy that must be designed in such a way that it elicits pattern recognition without causing overload.

The *semantically equivalent triplet* constitutes the smallest unit of this material. Each triplet consists of three sentences with identical semantic content — one in Spanish, one in Italian and one in Portuguese — and is based on the same lemma. The sentences differ in sound structure, morphological form and syntactic organisation, while remaining semantically equivalent. The strength of this principle lies in the fact that the learner does not learn three languages separately, but observes three representations of a single conceptual core. Within this controlled micro-variation, pattern recognition arises naturally: the brain compares, abstracts and generalises.

Within this learning design, variation is strictly limited to a small number of grammatical categories essential for A1/A2 processing. For verbs, variation is restricted to conjugation by person and number and to a minimal set of tense forms. For nouns, only singular and plural are varied. Variation in grammatical gender and adjectives is not designed independently, but follows necessarily from the selected nominal forms. More complex grammatical phenomena, syntactic rearrangement and combined axes of variation are deliberately excluded. A maximum of one grammatical variation is activated per practice sentence, with the aim of enabling inductive pattern recognition without explicit grammar instruction.

Within this structure, *cloze sentences* are used: deliberately constructed sentences in which one lexical or grammatical unit is omitted. The learner completes this element on the basis of context, semantic expectation and form recognition. In this design, cloze sentences do not function as assessment instruments, but as instruments for pattern reinforcement: by actively reconstructing the missing form, the learner consolidates internal representations of lexical, morphological and syntactic relationships. The exercise thereby mirrors the mechanism of *predictive processing*: the brain predicts the most probable form and in doing so anchors the underlying patterns.

Two types of cloze sentence perform distinct functions. Trilingual cloze sentences support lexical development: for each lemma, three parallel sentences are constructed, each with one word omitted. By reconstructing the missing element in three languages simultaneously, the learner activates the lexical core of the system and anchors it in three contexts. Monolingual cloze sentences, by contrast, serve contextual awareness. They appear only after the lemma has become automatised and explore variation in meaning and use of already familiar words within the individual languages. The distinction between these two levels — trilingual for formation, monolingual for application — ensures a natural progression from cross-linguistic recognition to language-specific refinement.

The use of cloze exercises required a tightly bounded form of context. For this reason, the design uses *micro-contexts*: functional situations that evoke precisely one action or state and permit only one plausible lexical completion. These contexts are not intended as realistic or rich language situations, but as semantic anchors. They must not depend on implicit world knowledge, pragmatic conventions or cultural assumptions, and must remain realisable in all target languages without a shift in meaning. Micro-contexts thus function as a necessary condition for semantic stability, interlingual comparability and controllability of the practice materials.

The role of artificial intelligence within this process is strictly supportive. AI systems are used to apply variation consistently within the defined axes, not to generate arbitrary input. One model produces an initial set of candidates that satisfy the predefined semantic and grammatical parameters; a second control model evaluates whether the output reproduces the intended variation without losing the semantic core. Final selection remains a human responsibility, since naturalness, idiomatic usage and cultural appropriateness cannot be guaranteed automatically. This hybrid procedure ensures a degree of internal coherence in the material that no single model can autonomously guarantee.

The **design grammar** of the material can therefore be summarised as a balance between freedom and constraint. Within each triplet, variation is deliberately staged to make patterns visible; across triplets, controlled distribution provides repetition without redundancy. The cognitive logic underlying this approach is the same as that of *distributional learning*: meaning

emerges from patterned regularity within controlled diversity. By designing variation as a structural property rather than as incidental fluctuation, the grammar of the learning design is not written in rules but in recognised patterns. The material itself *is* the grammar.

§3.6 Micro-Curriculum and Scaffolding

The learning design has a layered temporal structure in which one crucial week — the *micro-curriculum week*, or *Week 0* — functions as the cognitive starting mechanism for the entire programme. During this initial phase, knowledge is not taught in the conventional sense; instead, the learner's perceptual capacities are systematically prepared for pattern recognition and interlingual comparison. The micro-curriculum therefore has a preparatory rather than productive function: it forms the bridge between the theoretical model of emergent learning and the practical implementation of the twelve-week programme.

Purpose and character of Week 0. Week 0 is designed as an intensive receptive week in which the learner is immersed in a controlled abundance of lexical and phonological input. Its purpose is to activate *distributional sensitivity*: the ability to perceive similarities and differences between related languages without yet having to name them. The learner practises looking, listening and comparing rather than producing language. In cognitive terms, this is the phase in which the *perceptual filter* is calibrated: the brain learns to distinguish relevant similarities from incidental formal variation (Ellis, 2003; Kuhl, 2004).

Structure of the micro-curriculum. The week consists of five days, each containing four blocks organised around a specific semantic domain, such as family, time, colour, emotions or objects. Each block contains thirty lemmas presented in parallel across the three languages. Week 0 therefore comprises twenty blocks and six hundred lemmas, distributed across twenty thematic clusters. Activities within each block follow a fixed sequence: (1) **Lexical presentation** — the learner sees and hears the words in Spanish, Italian and Portuguese, each in a short example sentence; (2) **Interlingual comparison** — the three forms are placed side by side, after which the learner marks similarities at the morphological, phonetic and graphic levels (strategies M, F and G); (3) **Cloze reconstruction** — one word is omitted from each of the three parallel sentences, and the learner supplies the missing element on the basis of context and visual similarity; (4) **Reflective observation** — the learner records which patterns or regularities have been noticed, such as corresponding suffixes or sound shifts. The activities are deliberately repetitive, but not redundant. By working through four semantic clusters in a single day, a high frequency of comparative observations is created. The learner develops an initial *interlingual sensitivity*: the ability to perceive structural correspondences without explicit metalanguage.

Cognitive function of the micro-curriculum. The core function of Week 0 is phonological and morphological priming. By hearing and seeing hundreds of parallel words and short sentences, the learner develops preliminary representations of sound and form patterns that later function as recognition anchors. Rather than learning rules, the learner develops *expectations* about how words in the three languages typically behave. This phase is comparable to the pre-linguistic *acclimatisation phase* observed in natural language acquisition (Bybee, 2010; Tomasello, 2008). The micro-curriculum also introduces the six core strategies (M, F, G, S, T and D) in their simplest form. Learners become familiar with how morphological and phonological patterns manifest themselves across the three languages. The week concludes with

a brief self-reflection in which learners identify their own observations: not for assessment purposes, but to develop awareness that pattern recognition is an active process.

Phasing of the main programme (Weeks 1–12). After Week 0, the main learning programme begins. It is divided into three broader phases marking the gradual transition from external support to internal regulation. **Phase 1 (Weeks 1–4)** focuses on lexical and phonological expansion. Learners extend the basic lexicon and recognise familiar patterns in new contexts. Activities remain predominantly receptive. **Phase 2 (Weeks 5–8)** focuses on syntactic and semantic consolidation. Axes of variation are used more systematically, while reflection questions develop from closed to semi-open formats. **Phase 3 (Weeks 9–12)** focuses on pragmatic integration. The emphasis shifts from recognition to application, and scaffolding becomes largely internal. This three-part structure reflects the temporal dynamics of the learning design: repetition at increasing levels of complexity. Each phase introduces new combinations of variation axes while retaining the pattern-based logic established in Week 0.

Scaffolding: from external to internal regulation. Within this design, the concept of *scaffolding* has a dual meaning: it refers both to instructional support for the learner and to the internal structure of the material itself. In the initial phase, scaffolding is predominantly external: AI systems provide the variation, Anki organises repetition, and reflection formats direct attention. As the course progresses, the emphasis shifts towards internal scaffolding: learners use previously acquired patterns as mental supports when interpreting new input. By Week 12, the degree of external support is minimal, and the system functions as a self-regulating learning network.

The temporal logic of scaffolding is cyclical: each week contains micro-repetitions of the sequence *input → observation → reconstruction → reflection → repetition*. Learning thereby becomes recursive rather than linear, increasing the durability of acquired patterns.

Role of the instrumentation. The instrumentation of the learning design does not constitute a separate instructional domain, but an integrated functional infrastructure in which human and artificial components jointly shape the learning environment. Each instrument performs a specific task within the scaffolding cycle, orchestrating the interaction between exposure, repetition, pronunciation and reflection. Together, these tools ensure that the grammar-free learning process remains systematic, repeatable and empirically verifiable.

The first instrument is the AI pair **ChatGPT** and **Claude**, used within a separate and hierarchically organised production and control process. Practice materials are generated with ChatGPT under strict direction from the researcher and within predefined micro-batches and micro-contexts. This direction includes lemma selection, context construction, monitoring of semantic equivalence and organisation of variation. Claude is not used in this process as an alternative generator or as a substantive evaluator, but exclusively as an external control instrument for an initial verification of translation accuracy between the target languages. This control focuses explicitly on whether the derived sentences in the other languages constitute correct translations of the reference sentence and does not include reassessment of context, instructional suitability or interlingual design choices. This limited and clearly bounded use prevents the verification process from acquiring the same level of complexity and labour intensity as the generative micro-process itself. Claude's role is therefore supplementary and technical, while substantive and instructional responsibility remains entirely with the human curator.

The second instrument, **Anki**, translates the theoretical principles of frequency and distribution into an operational memory mechanism. Its spaced-repetition algorithm determines, on the basis of performance and response time, how often each lemma recurs. This creates an individual rhythm of exposure in which strongly consolidated patterns recur less frequently while weaker patterns recur more often. Anki therefore functions as the **temporal engine** of the learning design: it makes abstract principles such as frequency and distribution operational and gives the learning process a data-driven cadence.

A third, often underestimated component is **Speechling**, the platform that supports the auditory dimension of the learning design. Whereas ChatGPT and Claude provide textual input and Anki structures memory practice, Speechling provides an environment for pronunciation and listening practice that activates the phonological component of scaffolding. Learners hear native speakers of the three target languages, repeat the sentences and receive immediate feedback on pronunciation accuracy. In this way, the phonological form of words is linked to their visual and semantic representation. This auditory repetition reinforces the perception of sound and rhythm patterns introduced receptively during Week 0. Speechling thus functions as auditory scaffolding: it stabilises the sound–form–meaning triangle on which further pattern formation rests.

The fourth instrument consists of the **reflection formats**, which provide the metacognitive completion of the system. They support the transition from implicit recognition to conscious observation. After each block or week, learners answer brief questions focused on observation rather than explanation: Which forms occurred more frequently? Which language uses shorter words for the same concept? Such targeted observation tasks develop metalinguistic awareness without explicit grammar instruction. The reflection formats therefore function as the cognitive link between the artificial consistency of the AI layer and human interpretation of meaning.

Together, these four instruments create a *hybrid learning environment* in which artificial precision and human interpretation complement one another. ChatGPT and Claude provide controlled variation, Anki manages temporal distribution, Speechling anchors the auditory dimension, and the reflection formats provide conceptual anchoring. Their interaction operationalises the principle that learning without grammar is not the same as learning without structure: structure emerges here not from rules, but from the coordinated interaction of technology, repetition and reflection. This infrastructure makes grammar-free acquisition not accidental, but *methodologically controllable*: each instrument records data that can later be used to evaluate and verify the learning mechanisms. In this sense, the instrumentation is not merely supportive, but constitutive of the scientific character of the learning design.

§3.7 Expected Learning Mechanisms and Testability

The learning design functions not only as an instructional model, but also as a research instrument through which hypotheses concerning language acquisition in a multilingual, grammar-free context can be tested. The instrumentation described in the preceding section — ChatGPT/Claude, Anki, Speechling and the reflection formats — creates a rich ecosystem in which cognitive processes can not only emerge but also be recorded. Each instrument produces data that provide insight into the dynamics of learning: frequency curves, error patterns,

pronunciation evaluations and reflective notes together form a multidimensional representation of the learning process.

The central hypothesis of this research is that simultaneous exposure to three related languages under conditions of controlled variation leads to **emergent grammatical sensitivity**: the learner's ability to recognise structural similarities and differences without explicit instruction. This process is expected to unfold through a series of cognitive mechanisms corresponding to the phasing of the micro-curriculum and the subsequent twelve weeks.

Phase 1: Perceptual pre-activation (Week 0). During Week 0, the micro-curriculum, the first pre-linguistic learning mechanisms are activated. Through intensive exposure to trilingual lemmas, the learner develops increased sensitivity to morphological and phonological patterns. This phase stimulates what is referred to in the cognitive literature as *statistical learning* (Saffran et al., 1996): the ability to detect frequent combinations of forms automatically. Cloze exercises and interlingual comparisons reinforce the perception of distributions, while Speechling supports the phonological formation of auditory categories. Testability in this phase is based on perceptual recognition tasks: the speed and accuracy with which learners identify recurring patterns are recorded as indirect measures of emergent sensitivity.

Phase 2: Implicit generalisation (Weeks 1–4). In the first phase of the main programme, the emphasis shifts from perception to the formation of generalisations. Learners begin implicitly to recognise clusters of related forms as variants of a single underlying schema. In the literature, this process is described as *exemplar-based abstraction* (Bybee, 2010): new forms are not stored separately, but integrated into existing patterns. The triplet structure facilitates this through repetition under controlled variation, while Anki provides temporal consolidation. Testability here is based on frequency analysis of corrections in cloze tasks: a declining error rate across comparable lemmas indicates strengthening of the underlying representations.

Phase 3: Structural convergence (Weeks 5–8). During the consolidation phase, learners are expected to begin recognising structural correspondences that extend beyond the boundaries of individual languages. The observation of parallels in word order, negation or aspect provides the basis for *syntactic alignment* (Pickering & Garrod, 2013). This convergence is reinforced through reflection formats in which learners explicitly indicate which sentence components they believe to be functionally equivalent. AI-driven variation provides sufficient contrasts to make the phenomenon measurable: the frequency with which learners identify correct alignments is recorded as an indicator of syntactic understanding without explicit grammar instruction.

Phase 4: Pragmatic integration (Weeks 9–12). The final phase of the learning programme activates the pragmatic component of language acquisition. Learners apply patterns within communicative situations that vary in formal register, intention and context. At this point, *transfer-appropriate processing* (Morris et al., 1977) becomes relevant: knowledge acquired during training in realistic contexts is activated more effectively during application. Monolingual cloze sentences and the pronunciation component of Speechling provide empirical data on context sensitivity: the extent to which learners select the appropriate form or intonation for a given situation functions as a measure of pragmatic flexibility.

Interaction between instrumentation and cognitive mechanisms. The four instruments function as complementary observation points within a single learning process. *ChatGPT/Claude* document the consistency of input and can provide frequency profiles of the

structures used. *Anki* records precisely how many repetitions are required before an item is mastered, providing a direct measure of consolidation speed. *Speechling* supplies quantitative data on pronunciation accuracy and change over time. *Reflection formats* generate qualitative data on awareness and pattern recognition. Together, these layered data provide an empirical window onto the cognitive dynamics of grammar-free acquisition. They make it possible to analyse how emergent knowledge manifests itself in behaviour without directly measuring explicit knowledge content.

Internal validation and triangulation. The testability of the design rests on *internal triangulation* between three types of observation: behavioural data, such as response time and correction; auditory data, such as pronunciation feedback; and metacognitive data, such as reflection reports. Combining these sources makes it possible to establish not only *whether* learning occurs, but also *how* it develops. Convergence between patterns — for example, declining error frequency in Anki, improved pronunciation in Speechling and more accurate observations in the reflection formats — provides the empirical test of the central hypothesis: that grammar-free acquisition, when appropriately structured, leads to stable form–meaning representations across three languages. Week 0 plays a crucial role in this analysis. Because all learners receive the same initial input, the micro-curriculum functions as a *baseline measurement*: variation in later performance can be attributed to differences in depth of processing rather than unequal exposure. The entire system is designed so that every intervention — every repetition, pronunciation task or reflection — documents itself. The learning design thereby becomes a *self-recording experiment* in which the boundary between instruction and empirical research becomes less distinct.

Scientific implication. The learning design illustrates that grammar-free language acquisition need not remain elusive. Through the combination of controlled variation, technologically supported scaffolding and systematic data collection, emergent learning can be made observable. Testability does not reside in individual measurements, but in consistency of patterns across instruments: when different data streams — memory, pronunciation and reflection — point in the same direction, emergent learning becomes plausible. This creates a research model in which the learning process itself provides the evidence. The instrumentation, designed as instructional support, becomes a measurement network through which the validity of the underlying theory can be assessed. The result is not merely a description of a learning design, but a blueprint for an empirically testable hypothesis concerning multiple language acquisition without explicit grammar instruction.

§3.8 The Unpredictability of the Design Process

During implementation of the learning design, the production of suitable practice materials proved considerably less predictable and more difficult to control than could have been anticipated. This unpredictability did not concern the instructional premises themselves, but their practical translation into practice materials that could be produced at scale while remaining consistent and controllable. Precisely because the design is based on grammar-free, inductive and simultaneous polyglot learning, limitations and tensions emerged that are less readily visible in a traditional monolingual learning design.

3.8.1 The emergence of semantic drift. A first structural problem concerned not the meaning of individual sentences, but the cumulative effect of small deviations that accumulated during longer and larger-scale generation processes. When production took place in larger batches, language models were found gradually to move away from the original formulation of the task. This deviation did not manifest itself as an abrupt error, but as an accumulation of incremental shifts that became visible only when the material was considered as a whole. The result of this accumulation was increasing semantic flattening. Sentences remained grammatically correct and locally plausible, but gradually lost their functional precision. Particularly towards the end of production processes, the proportion of generic formulations increased and unwanted placeholders appeared, not as incidental errors but as symptoms of diluted adherence to the original instruction. The output remained formally acceptable, but increasingly diverged from the intended instructional precision.

This form of semantic drift is difficult to detect automatically precisely because it rarely results in binary errors. Individual sentences provide little reason for correction; the problem becomes apparent only at scale. At its core lies a structural limitation of language models: they have a limited instruction horizon. As a generation process becomes longer and encompasses more output, the probability increases that previously imposed constraints will be applied less strictly. The difficulty therefore is not incorrect output, but in the model's inability to maintain semantic precision consistently over extended production. The design implication is that semantic stability cannot be imposed retrospectively, but must be protected in advance. This requires actively bounding production processes, dividing them into small, controllable units and anticipating cumulative deviations. Semantic drift is therefore not an incidental model failure, but a systemic effect that can only be managed through design interventions.

3.8.2 Problems in designing micro-contexts. Closely related to semantic drift were the difficulties surrounding micro-contexts. Within this learning design, contexts are not intended to depict rich or realistic situations, but to constrain meaning so precisely that only one plausible completion remains possible within a cloze structure. In practice, this proved to be a delicate balance. Contexts that were too open permitted multiple interpretations. A sentence such as *"He ____ the door"* may evoke *opens*, *closes*, *paints* or *repairs*. Such contexts undermined the inductive function of the exercise. At the same time, contexts that were too sparse or artificial provided insufficient guidance, causing sentences to lose their linguistic credibility. Additional complexity arose because contexts also had to remain realisable interlingually. A context that functions unambiguously in Spanish may inadvertently activate a different interpretation in Italian or Portuguese. Designing micro-contexts therefore proved not to be a trivial preparatory step, but an iterative process in which semantic precision was continually under pressure.

3.8.3 The emergence of grammatical nonsense. A difficulty emerged in what is referred to here as grammatical nonsense: sentences that are formally correct but functionally or semantically empty or incoherent. An example would be a sentence such as *"I cook mirror."* Syntactically, the structure is recognisable — subject, verb, object — but the combination of verb and object is meaningless. In monolingual contexts, this is readily apparent; in large-scale multilingual generation, however, such nonsense can emerge unnoticed, particularly when axes of variation are combined without sufficient semantic anchoring. The problem here is not grammatical incorrectness, but the absence of *meaningful selectional restrictions*. For inductive learning, grammatical correctness is insufficient; sentences must also be functionally interpretable. It was precisely this layer that proved vulnerable in automatic generation.

3.8.4 Problems in creating semantic triplets. The ambition to present practice materials trilingually introduced an additional source of difficulties. Creating *semantically equivalent sentences* in three languages simultaneously proved fundamentally more problematic than generating separate monolingual sentences. When sentences are generated in parallel, there is a risk that equivalence is assumed rather than verified. Small shifts in aspect, temporal nuance or pragmatic focus can result in sentences that are formally comparable but functionally divergent. This is particularly problematic when the sentences are intended as comparative material for inductive pattern recognition. The difficulty lay not in translation errors, but in the fact that semantic equivalence is not an automatic by-product of correct sentence generation. It requires explicit control.

3.8.5 Displaying understanding without awareness. An intriguing aspect of the production process was that, despite lacking awareness or intention, the language model proved capable of formulating adequate reflections and corrective proposals once a human curator identified an error. The model did not autonomously 'recognise' errors, but once they had been pointed out, it could provide coherent explanations and suggest alternatives. This phenomenon can easily be misinterpreted as understanding. In reality, it reflects pattern sensitivity at a meta-level: the model can identify inconsistent or problematic structures when they are explicitly indicated, but does not necessarily identify them in advance. This asymmetry between generation and correction contributed to the unpredictability of the process.

3.8.6 Scale and the threat of combinatorial explosion. The combination of several languages, several parts of speech and several axes of variation led to a rapid increase in the number of possible sentences and variants. This combinatorial explosion was not merely a theoretical possibility: it posed a genuine practical threat to manageability. The problem was not the generation of individual sentences, but maintaining an overview, coherence and semantic consistency across larger bodies of material. Without explicit constraints, the design risked undermining itself: not because it became incorrect, but because control would become practically impossible

3.8.7 Human control without language proficiency. Finally, implementation exposed a fundamental tension in the role of human control. The human curator does not possess productive proficiency in the target languages, making comprehensive linguistic assessment impossible. At the same time, control proved not only possible but necessary. Crucially, control was not exercised on the basis of native-like intuition, but on structure, consistency of meaning and interlingual parallelism. Deviations became visible when contexts ceased to be unambiguous, sentences lost their function as exercises, or parallels between languages broke down. Control thereby shifted from language proficiency to understanding of the design.

3.8.8 Interpretation. The forms of instability and unpredictability described here could not have been fully anticipated in advance and could not simply have been prevented by refining the design principles. They are inherent in the combination of scale, inductive ambition and multilingual parallelisation. They therefore do not constitute a refutation of the learning design, but an empirical boundary within which it became feasible. The feasibility of grammar-free, inductive and polyglot learning thus proves to depend not only on instructional principles, but to a substantial extent on the manageability of the production process itself.

§3.9 From Unpredictability to Manageability

The difficulties described in §3.8 did not lead to a revision of the design principles, but to a series of concrete and enforceable design measures that made the production of practice materials manageable. First, the generative space was strictly divided into micro-batches: small, self-contained production steps in which only a limited number of lemmas were developed at a time. This choice functioned as the primary measure against semantic drift and enabled interim control before scaling up. Within each micro-batch, strictly defined micro-contexts were used as stable semantic anchors. These contexts were not intended to be rich or realistic, but functional: they had to evoke one concrete, conceivable situation, activate one central action or state, and make no appeal to implicit world knowledge, pragmatic conventions or contextual enrichment. Crucially, each context was designed so precisely that it supported only one plausible completion within a cloze structure and could also be realised in all target languages without semantic reinterpretation. Contextual variation was therefore not eliminated, but subordinated to semantic stability and interlingual comparability.

For the tri-lingual construction of the practice materials, a serial derivation structure was adopted. Each set began with a single reference sentence in one language, which functioned as a semantic anchor; the corresponding sentences in the other languages were derived from it rather than generated in parallel. This prevented semantic equivalence from being implicitly assumed rather than explicitly controlled. Variation was likewise not designed trilingually from the outset, but was explicitly preceded by monolingual variation design in one source language (Spanish, ed.). In this language, the relevant patterns of variation were established for each part of speech, with only one axis of variation activated at a time. The sentences constructed in this way then served as a stable basis for interlingual derivation, with the other target languages inheriting this variation rather than reproducing it independently. This choice implies that language-internal possibilities for variation in the derived languages were not exploited to their full extent. Italian and Portuguese in particular possess their own potential for variation, which is addressed only selectively within this configuration. This limitation was deliberately accepted because asymmetric variation would undermine interlingual comparability and substantially increase the risk of semantic drift. Within the objective of controlled induction and stable trilingual comparison, symmetrical inherited variation proved more functional than language-internal completeness. Variation was furthermore not left unrestricted, but organised along predefined axes, with explicit stop rules and variation budgets determining when further expansion no longer offered instructional value and instead became counterproductive.

These measures did not make human control redundant, but did make it manageable. Control focused not on comprehensive grammatical assessment, but on consistency of meaning, contextual plausibility and mutual parallelism. Precisely for this reason, systematic control proved possible without active proficiency in the target languages: deviations became visible at the structural and semantic level rather than through linguistic intuition. The result is not a simplified design, but a stabilised implementation: the original principles of grammar-free, inductive and polyglot learning were retained, but embedded within a production process that proved feasible, albeit at the cost of considerable human effort. Chapter 3 therefore ends not with a promise, but with a controlled form of feasibility.

Chapter 4 – Research Methodology

§4.1 Introduction

This chapter sets out the methodological foundation of the research into AI-facilitated polyglot language acquisition. It concerns a practice-oriented study conducted from an autoethnographic perspective: the researcher is simultaneously the language learner, a reflective actor and the designer of the intervention programme. Rather than adopting a conventional design cycle involving complete iteration, a model was selected that more closely reflects the learning reality of an individual learner.

At the centre of the study is a *reflective-instructional research model in* which the learning programme itself functions both as the object of research and as a source of data. Reflection on the researcher's own learning process serves not merely as description, but is systematically employed as a method of analysis. This chapter therefore sets out how data are collected and analysed in order to answer the three empirical research questions. The learning book *Polyglottism & Linguistics* (version v03) provides the concrete implementation framework: all measurements, reflections and tasks are embedded within its weekly structure and generate the systematic data streams accounted for in this chapter.

§4.2 Research Model

This research lies at the intersection of language-acquisition research, language pedagogy and AI applications. Rather than conducting a controlled experiment involving multiple language learners, an autoethnographic design study with $n = 1$ was selected. The researcher is simultaneously language learner, designer and documenter. The study is therefore personal, because it is based on the learning experience of a single learner, and systematic, because that experience is recorded using fixed formats, linked to theoretical concepts and directly aligned with the three empirical research questions.

The research model employed can be characterised as a ***reflective-instructional framework***. This means that: 1) the learning programme itself constitutes the primary object of research; 2) reflections and weekly assignments constitute the data sources; and 3) the analysis focuses on the interaction between input (AI tools, learning book), processing (reflection, practice, comparison) and output (language production, error analysis).

The learning book *Polyglottism & Linguistics* functions as the organising framework: it provides the weekly structure, tasks and reflection formats through which data collection proceeds in parallel with the learning process. This approach is consistent with Design-Based Research (DBR), in which interventions are designed, implemented and studied concurrently, with the aim of contributing both to theory development and to practical insight.

§4.3 Methodological Positioning

The research is based on a design-oriented logic in which learning and measurement coincide. The learning programme is not merely a context for observation, but a controlled environment in which the cognitive operation of grammar-free language acquisition can become empirically visible. The methodological positioning of this research rests on two interconnected pillars: the combination of emergent and intentional forms of data, and the learning design as an integrated measurement instrument.

Methodological consequences of emergence and intentionality. The learning design combines implicit and explicit learning within a single integrated process. This choice has direct methodological consequences: it determines how learning processes are observed and which forms of data can count as evidence. Because the design stimulates emergent learning — knowledge that develops through exposure and frequency — most learning processes cannot be investigated directly through self-report. The researcher must therefore rely on indirect indicators: performance curves in Anki, success rates in cloze reconstruction, patterns in pronunciation feedback and improvement over time. These behavioural traces provide an empirical window onto processes that largely occur outside conscious awareness. At the same time, the design generates intentional data: observations, reflective responses and self-descriptions arising from the metalinguistic reflection deliberately incorporated into the programme. The analysis therefore encompasses not only behaviour, but also the meanings that language learners — or, in this case, the researcher-learner — themselves attribute to their learning experiences.

The methodological core lies in combining these two layers of data. Emergence provides quantitative indications of implicit development; intentionality provides qualitative clarification of what that development means cognitively. Integrating the two produces a hybrid evidential system that assigns equal importance to statistical patterns and introspective accounts. The distinction between emergent and intentional learning is therefore not merely theoretical, but methodologically functional: it determines the nature of the observations, the interpretation of the data and the criteria for validity. The result is a research strategy in which what emerges implicitly and what is articulated explicitly do not exclude one another, but jointly delimit the testable domain of grammar-free acquisition.

The learning design as a measurement instrument. Within this research, the learning design performs a dual function: it is both an instructional model and a research instrument. The learning book (version 3) is not merely the context in which learning takes place, but the structured environment within which the operation of the learning process can be observed and measured. Each component — the trilingual cloze sentences, the axes of variation, the reflection formats and the associated tools — is designed to generate data on a specific aspect of language acquisition. The material itself therefore *is* the measurement instrument: a controlled system that makes its own operation observable.

This configuration renders traditional distinctions between intervention and observation unnecessary. Whereas conventional experiments *measure* learning after the event, this design records learning *during* the process itself. The combination of AI-generated input, spaced repetition through Anki, pronunciation feedback through Speechling and reflective output produces a fine-grained trace of behavioural and performance data. Because all language

learners — or, in the case of autoethnography, the researcher-learner — receive identical learning stimuli, the learning book functions as a standardised research environment in which differences in performance can be traced directly to differences in processing. This approach redefines the epistemological position of the research. Rather than treating the learning design as an instructional tool that happens to produce data, it is positioned as a *self-recording experiment*: an instructional system that generates its own empirical evidence. The design thereby acquires not only educational but also scientific validity: every learning activity is simultaneously an observation, every variation a measurement point and every repetition a data point within a reproducible pattern.

The methodological positioning described in this section provides the analytical framework for the empirical chapters that follow. It determines not only *what* is observed, but also *how* those observations acquire meaning within a design that integrates learning and measurement. Section 4.4 shows how this integrated approach generates concrete data on language acquisition without explicit grammar instruction, and how those data can provide empirical support for the theoretical assumptions developed in Chapters 2 and 3.

§4.4 Data and Reflection Instruments

Data collection in this research is fully aligned with the three empirical subquestions and follows the principle of targeted, integrated data collection. Each question corresponds to a specific cluster of data types, reflection formats and measurement points. Because the language learner is also the researcher, the learning process itself constitutes the primary source of observation and analysis. The data collected are neither separate from the learning process nor retrospective, but embedded within the weekly rhythm of the learning programme.

For each subquestion, the type of evidence considered valid was defined in advance. These evidential standards are linked to specified data sources in order to exclude alternative explanations, such as chance accuracy or mere exposure.

Subquestion 1 – A2 level in an accelerated, grammar-free programme with AI support
To what extent can A2 level be achieved in three related languages within an accelerated learning programme involving minimal grammar instruction and support from AI tools?

- **Evidence requirement:** Consistent performance on CEFR-A2-based output tasks in unfamiliar contexts that cannot be explained by chance accuracy or pre-existing knowledge.
- **Required data:**
    - Scores on CEFR-based control tasks (e.g. introducing oneself, providing a simple description).
    - Anki retention data demonstrating that lexical knowledge has been retained over time (e.g. consistent recall of 75% of words after at least three weeks).
    - Speechling recordings in which intelligibility is confirmed by a native speaker (‘sufficiently intelligible without repetition’).
    - Process data from weekly tasks demonstrating gradual development (progression in output complexity).
- **Safeguarding internal validity:**
    - A baseline measurement (T0) confirms the absence of relevant prior knowledge.

    - CEFR tasks use new contexts to exclude memorisation of practised examples.
    - No feedback is provided during summative tasks, in order to prevent learning effects during testing.
    - Comparison with CEFR descriptors ensures objective assessment.

Subquestion 2 – Metalinguistic awareness and pattern recognition: How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme?

- **Evidence requirement:** There must be evidence that the language learner can consciously identify observed patterns and apply them correctly in new tasks in a way that cannot be explained by chance recognition or repeated exposure alone.
- **Required data:**
    - Logbook entries explicitly identifying characteristics of patterns.
    - Annotations accompanying AI interactions in which choices are justified.
    - Cloze exercises in which missing words are supplied and the language learner explains the choice. (Cloze exercises are fill-in sentences in which one element is omitted, intended to stimulate active pattern recognition and contextual recall.)
    - Correct application of articulated patterns in new production tasks.
- **Safeguarding internal validity:**
    - Reflections are linked to specific tasks.
    - Patterns are presented in different variants; correct application must occur in a new variant.
    - The sequence is recorded: explicit identification → application.
    - Where possible, control tasks without reflection are used to test whether progress can be attributed to reflection.

Subquestion 3 – Relatedness as an accelerating factor: To what extent does structural relatedness between the target languages function as an accelerating factor in simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?

- **Evidence requirement:** Relatedness must demonstrably result in acceleration — faster task performance and fewer errors — that cannot be explained by general improvement in proficiency.
- **Required data:**
    - Time measurements and error rates for tasks involving related versus unrelated structures.
    - Logbook entries in which relatedness is explicitly identified.
    - Production tasks with controlled variation (Spanish–Italian versus Spanish–German).
- **Safeguarding internal validity:**
    - Tasks with balanced levels of difficulty and complexity.
    - Evidence is accepted only when relatedness is explicitly identified.
    - Control tasks without relatedness are used to exclude general proficiency effects.
    - Standardised conditions are used for time measurements.

Validation of AI-generated learning materials. Because a substantial proportion of the learning materials — example sentences, parallel translations, cloze exercises and reflection

prompts — was generated using generative AI, their validation is methodologically crucial. A systematic control procedure was therefore incorporated:

- Claude was used to generate semantically identical triplets.
- ChatGPT was used for verification of grammatical correctness and semantic parallelism.
- An independent linguist conducted sample-based checks for:
  - lexical correctness,
  - grammatical correctness,
  - semantic equivalence.
- Where systematic errors were identified, the batch was regenerated or modified.

This procedure strengthens the internal validity of the findings by excluding the possibility that observations are based on incorrect or misleading input. The procedure is documented in the project logs and illustrated in the learning book (see Weeks 1–2 for examples of cloze triplets).

§4.5 Validity, Reliability and Transparency

The reflective nature of this research — in which the language learner is also the researcher — requires particular attention to validity, reliability and transparency. These dimensions are essential to ensure that the findings can be scrutinised and that alternative explanations can be excluded.

1. Internal validity: comparison between initial and final proficiency. A baseline measurement (T0) in Week 0 establishes the initial proficiency level across four dimensions: receptive vocabulary, functional sentence production, pronunciation recognition and basic understanding of language-use situations. In Week 12, a final measurement is administered using identical or parallel formats. This comparison makes it possible to establish whether the intended A2 level has been reached — not only for each language separately, but also in terms of overall communicative proficiency.

2. Structural reliability of reflection data. The logbook data are not recorded in free-form format, but collected through fixed weekly formats covering three domains of reflection:

- *What did I notice?* (observation),
- *Why did this happen?* (explanation),
- *How did I respond?* (regulation).

In addition, reflections are structured using semantic markers (e.g. “I confused”, “I discovered”, “I recognised”, “I adjusted”). This increases consistency and enables systematic analysis. The learning book (v03) contains the relevant formats and tasks, ensuring that the data collected are closely aligned with the research questions.

3. Triangulation between data streams. To validate subjective impressions, three types of data are consistently combined:

- logbook reflections,

- AI output or feedback (e.g. corrections, paraphrases),
- system data from Anki and Speechling.

For example, when the language learner records in the logbook, "I suddenly understood this construction", it is examined whether that construction is also used correctly in ChatGPT or Claude output and whether it recurs in Anki. This creates either mutual corroboration or tension between sources, thereby strengthening validity.

4. Transparency and verifiability. All data used — prompts, responses, feedback and reflections — are stored with timestamps and version numbers. Reflections are not modified retrospectively. Micro-iteration reports record which changes were made, why they were made and what effect was intended. In addition, an explicit evidential standard applies to each phenomenon: an observation is accepted as evidence only when it meets predefined criteria, such as the application of an identified pattern in a new context or consistently correct production in unfamiliar situations. This excludes chance accuracy, memorisation or general proficiency development as alternative explanations for the findings.

§4.6 Evaluation Strategy for Each Research Question

For each empirical subquestion, the collected data are evaluated using predefined analytical methods and assessment criteria. The evaluation follows directly from the formats and instruments embedded in the learning book (v03), including Anki retention data, Speechling recordings, cloze exercises, logbook entries and AI interactions.

Subquestion 1 – A2 level in three related languages with AI support: *To what extent can A2 level be achieved in three related languages within an accelerated learning programme involving minimal grammar instruction and support from AI tools?*

- **Basis for assessment:**
  - Results from T0 and T12: vocabulary, sentence construction and communicative performance.
  - Anki retention and repetition frequency.
  - Speechling recordings with external assessment by native speakers.
  - Self-assessment of functional language proficiency for each language.
- **Method of analysis:**
  - Qualitative comparison of T0 and T12 writing tasks.
  - Quantitative analysis of word recognition and automatisation in Anki.
  - Thematic coding of logbook reflections on communicative functioning.
- **Assessment criteria:**
  - Achievement of CEFR A2 objectives for each language.
  - Improvement between T0 and T12 in vocabulary and sentence production.
  - External confirmation of intelligibility.

Subquestion 2 – Metalinguistic awareness and pattern recognition: *How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme?*

- **Basis for assessment:**

    - Logbook entries explicitly identifying patterns.
    - Annotations accompanying AI interactions.
    - Cloze exercises with explanations of the choices made.
- **Method of analysis:**
    - Open coding of reflections on pattern recognition.
    - Axial coding of relationships between reflections and applied knowledge.
    - Thematic aggregation into broader learning mechanisms.
- **Assessment criteria:**
    - Demonstrable identification **and** application of patterns.
    - Correct use of patterns in new contexts.
    - Increase in metalinguistic awareness over the course of the programme.

Subquestion 3 – Relatedness as an accelerating factor: *To what extent does structural relatedness between the target languages function as an accelerating factor in simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?*

- **Basis for assessment:**
    - Comparative production tasks involving related versus unrelated structures.
    - Time measurements and error rates.
    - Logbook entries in which relatedness is explicitly identified as a learning aid.
- **Method of analysis:**
    - Thematic coding of reflections on transfer recognition.
    - Quantitative analysis of speed and accuracy.
    - Comparison between related and unrelated tasks.
- **Assessment criteria:**
    - Faster and more accurate performance on related structures.
    - Explicit reference to relatedness as a learning aid.
    - Demonstrable improvement relative to tasks without relatedness.

The evaluation of the individual research questions thus provides an initial picture of the extent to which functional proficiency, metalinguistic awareness and language relatedness as an accelerating factor contribute to the intended learning outcome. In the following section (§4.7), these evaluations are brought together in an integrated data analysis linking the three subquestions to the central research question.

§4.7 Data Analysis

The analysis of the collected data is primarily directed towards answering the central research question: *under what conditions can AI-facilitated support contribute to accelerating the simultaneous acquisition of three related languages to CEFR level A2, without loss of grammatical understanding or functional proficiency?*

This central question is operationalised through the three empirical subquestions set out in §§4.4–4.6. The analysis is therefore structured around three complementary pillars: functional proficiency (subquestion 1), metalinguistic awareness (subquestion 2), and relatedness as an accelerating factor (subquestion 3).

1. Functional proficiency (subquestion 1). The analysis of functional language proficiency focuses on demonstrating CEFR A2 level in the three target languages. Three data sources are integrated for this purpose:

- performance on CEFR-based communicative tasks;
- retention data from Anki (vocabulary development and automatisation);
- assessment of pronunciation and intelligibility through Speechling.

The qualitative comparison of T0 and T12 output is supplemented by quantitative analyses of word recognition and retention rates. Together, these provide evidence as to whether the required threshold of functional proficiency has been reached.

2. Metalinguistic awareness (subquestion 2). To obtain insight into the role of reflection and pattern recognition, thematic coding is applied to reflections and logbook data. Coding schemes are derived from the theoretical frameworks developed in Chapter 2 and the learning design described in Chapter 3, with categories such as *transfer recognition, error awareness, pattern comparison* and *interlingual strategy*.

The focus is on moments when the language learner explicitly articulates *why* a particular choice was made, and on whether that insight is subsequently applied correctly in new contexts. This makes it possible to distinguish incidental accuracy from structural understanding.

3. Relatedness as an accelerating factor (subquestion 3). To test whether structural relatedness leads to acceleration, comparative tasks are analysed in which the same meaning is produced in three languages. The analysis considers:

- task duration and error rates;
- explicit reflections in which relatedness is identified as a learning aid;
- differences between related and unrelated structures in controlled tasks.

Triangulation of task results, logbook reflections and output analyses makes it possible to determine whether gains in speed can genuinely be attributed to language relatedness, or whether alternative explanations, such as a general increase in language proficiency, are more plausible.

Synthesis. The three pillars are integrated within an iterative analytical cycle: qualitative data — logbooks, ChatGPT interactions and reflections — are coded and clustered, while quantitative data — scores, retention rates and time measurements — are used to objectify the patterns identified. This triangulation increases reliability and produces a rich corpus directly aligned with the central research question.

The purpose of this combined approach is not only to assess the outcome — attainment of A2 level — but also to identify the mechanisms that may make this possible: reflection, transfer and AI-facilitated support. Data analysis thereby provides the methodological bridge between design (Chapter 3), implementation (Chapter 4) and evaluation (Chapters 5–6).

§4.8 Methodological Reflection

This study inevitably has methodological limitations. It is an $n = 1$ study, which limits generalisability. This is not, however, an incidental limitation, but a deliberate choice of

research design in which *in-depth observation and triangulation* take precedence over broad representativeness. Evidential strength is derived from consistent performance on CEFR tasks, logbook reflections and system data from Anki and Speechling, which together provide a valid representation of the learning process.

A second consideration concerns the use of AI. The materials used were generated in part with Claude (lexicon and idiom), ChatGPT (writing support and reflection), Anki (memory training) and Speechling (pronunciation). This multiple use makes AI both a tool and an object of research. Although validation procedures have been incorporated, the reproducibility of individual outputs cannot be fully guaranteed. A *systematic documentation practice* was therefore adopted: all prompts, outputs and corrections were stored and version-coded, ensuring the traceability of decisions.

Third, the researcher is also the language learner. This dual role inevitably entails subjective involvement, but within this autoethnographic design it is deliberately used as a *functional source of data*. Through fixed coding schemes, triangulation and transparent reporting, subjectivity is not eliminated but systematically channelled.

Despite these limitations and design choices, the research provides valuable insights into how AI-supported polyglot learning can be designed, implemented and evaluated. The methodological choices are aligned with the object of research and thereby provide a consistent foundation for the empirical analysis.

§4.9 Conclusion

In Chapter 4, the methodological foundation of this research has been set out. The research design, the instruments and formats employed, and the procedures for data collection and analysis in conjunction with the learning book have been described. It has also been made explicit how validity, reliability and transparency are safeguarded despite the limitations associated with an $n = 1$ study and the use of AI.

The core of the approach lies in the integration of three elements: (1) a carefully designed learning programme directly linked to the learning book; (2) the use of AI tools as sources of input, feedback and reflection; and (3) an integrated data-collection plan that triangulates language production, reflection and system data. Through this combination, the central research question — under what conditions AI can accelerate the simultaneous acquisition of three related languages to A2 level — can be addressed empirically.

Chapter 5 – Empirical Testing

5.1 Introduction

This chapter presents the templates that will be used to report and analyse the empirical results of this research. The templates are derived directly from the evidential requirements and evaluation criteria established in Chapter 4 (§§4.4–4.6). They show which data are collected for each research question, how these data are organised, and which assessment standards are applied.

5.2 Research Question 1 – A2 Level in Three Languages

Introduction and evidential structure. According to §4.6, the first research question (*To what extent can A2 level be achieved in three related languages within an accelerated learning programme involving minimal grammar instruction and support from AI tools?*) may be answered positively if the following evidential requirements are met:

- *(X1)* the learner reaches CEFR A2 level in Spanish, Italian and Portuguese, operationalised through measure M1 = communicative tasks (T0/T12);
- *(X2)* demonstrable improvement occurs between T0 and T12 in vocabulary and sentence production, operationalised through measure M2 = Anki retention and repetition frequency + writing tasks;
- *(X3)* pronunciation and intelligibility are externally confirmed by native speakers, operationalised through measure M3 = Speechling recordings with external assessment;
- *(X4)* self-assessment of functional language proficiency supports the findings from the objective measurements, operationalised through measure M4 = self-report for each language.

The empirical data relating to these conditions are presented in the following tables. *(→ insert the tables for Subquestion 1 here)*

With regard to Research Question 1, it can be established that:

- condition X1 (CEFR A2 level) has [not / partially / fully] been met;
- condition X2 (lexical and syntactic improvement) has [not / partially / fully] been met;
- condition X3 (intelligibility confirmed by native-speaker assessment) has [not / partially / fully] been met;
- condition X4 (self-assessment of functional proficiency) has [not / partially / fully] been met.

On this basis, the research question as a whole can be answered [not / partially / fully] positively, in accordance with the assessment framework established in §4.6.

§5.3 Research Question 2 – Metalinguistic Awareness

Introduction and evidential structure. According to §4.6, the second research question (*How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme?*) may be answered positively if the following evidential requirements are met:

- *(Y1)* the learner can explicitly identify and describe linguistic patterns, operationalised through measure N1 = logbook entries containing explicit pattern recognition;
- *(Y2)* the learner can apply these patterns correctly in new contexts, operationalised through measure N2 = cloze exercises accompanied by explanations of the choices made;
- *(Y3)* the AI interactions demonstrate increasing metalinguistic depth, operationalised through measure N3 = annotations accompanying AI sessions and logbook reflections;
- *(Y4)* there is a demonstrable increase in metalinguistic awareness over the course of the programme, operationalised through measure N4 = coded reflections (Weeks 0–12).

(→ insert the tables for Subquestion 2 here)

Conclusion. With regard to Research Question 2, it can be established that:

- condition Y1 (**explicit identification of patterns**) has [not / partially / fully] been met;
- condition Y2 (**application in new contexts**) has [not / partially / fully] been met;
- condition Y3 (**metalinguistic development in AI interactions**) has [not / partially / fully] been met;
- condition Y4 (**increase in awareness over the course of the programme**) has [not / partially / fully] been met.

On this basis, the research question as a whole can be answered [not / partially / fully] positively, in accordance with the assessment framework established in §4.6.

§5.4 Research Question 3 – Relatedness as an Accelerating Factor

Introduction and evidential structure. According to §4.6, the third research question (*To what extent does structural relatedness between the target languages function as an accelerating factor in simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?*) may be answered positively if the following evidential requirements are met:

- *(Z1)* in tasks involving structural relatedness between the target languages, the learner shows evidence of acceleration or error reduction, operationalised through measure P1 = observations of time, accuracy and reflective explanation within comparative production tasks;
- *(Z2)* the reflections contain explicit references to relatedness as a learning aid, operationalised through measure P2 = logbook entries and annotations concerning transfer and interlingual comparison;

- *(Z3)* divergence cases function as an internal contrast condition: the learner demonstrates faster or more accurate processing of related structures than of divergent structures, operationalised through measure P3 = comparative analysis of tasks with and without D-marking (time measurement, errors and reflective explanation).

The empirical data relating to these conditions are presented in the following tables.

(→ insert the completion tables for Subquestion 3 from Chapter 5 here)

Conclusion. With regard to Research Question 3, it can be established that:

- condition Z1 (acceleration or error reduction in related structures) has [not / partially / fully] been met;
- condition Z2 (reflective reference to relatedness as a learning aid) has [not / partially / fully] been met;
- condition Z3 (contrast between related and divergent structures) has [not / partially / fully] been met.

On this basis, the research question as a whole can be answered [not / partially / fully] positively, in accordance with the assessment framework established in §4.6.

§5.5 Changes to the Approach – Micro-Iterations as Embedded Reflection Points

The learning programme was designed with three planned reflection points: after Week 4, Week 8 and at the end of the programme (Week 12). These points, also referred to as *micro-iterations*, are intended to test the design against practical experience and to adjust it where necessary.

A. Purpose of the iterative analysis

The purpose of this section is to describe how the findings from the three empirical subquestions jointly lead to:

1. identification of strong and weak components within the learning design (lexical development, scaffolding, interlingual strategies);
2. formulation of design adjustments on the basis of observed patterns;
3. assessment of the internal consistency between the design principles (Chapter 3) and the learning outcomes (Chapter 5).

B. Format for micro-iterative reporting

Each design iteration is reported according to the following fixed format:

Iteration phase	Empirical input	Observed effect	Design decision / adjustment	Supporting evidence (evidence IDs)
Weeks 0–4 (Iteration 1)	Placeholder: data from §§5.2–5.4			

Iteration phase	Empirical input	Observed effect	Design decision / adjustment	Supporting evidence (evidence IDs)
Weeks 5–8 (Iteration 2)	Placeholder: data from §§5.2–5.4			
Weeks 9–12 (Iteration 3)	Placeholder: data from §§5.2–5.4			

Completion instructions:

- *Empirical input* refers to specific findings from Chapter 5.
- *Observed effect* briefly records the implication for the learning process or materials.
- *Design decision* describes the adjustment to the weekly structure, lexicon, strategy or instrumentation.
- *Supporting evidence* lists the evidence IDs — Anki, Speechling or logbook files — supporting the decision.

C. Evaluation framework for iterations

Once the table has been completed, each iteration is briefly evaluated in terms of:

- whether the change strengthened or undermined the original design principle;
- whether additional empirical verification is required;
- which components of the design remained *stable* across iterations.

D. Placeholder for synthesis

After the three iterations, a summarising paragraph follows — to be completed after data processing — in which the iterative cycle is discussed in relation to Chapter 3:

"The cumulative adjustments show that [learning component X] improved consistently, whereas [component Y] requires further optimisation."

E. Transition to Chapter 6

The iterative analysis forms the link between empirical evidence and interpretation. Chapters 6 and 7 address how these design experiences are interpreted theoretically within the broader discussion of AI-facilitated polyglot language acquisition.

Chapter 6 – Conclusions, Discussion and Implications

§6.1 Introduction

The purpose of this chapter is to answer the three empirical research questions of this dissertation and to critically assess the validity of those answers. Whereas Chapter 4 established the methodological basis and Chapter 5 presented the results, the focus here shifts to interpretation: what can actually be concluded on the basis of the collected data and analyses, and which limitations must be explicitly acknowledged?

Three principles guide this analysis. First, the research questions are discussed separately, in each case in relation to the predefined evidential standards (see Chapter 4). Second, for each question, the strength of the empirical indications is weighed against possible alternative explanations. Third, an overarching reflection addresses the limits of generalisability, the role of AI-generated learning materials and the implications for further research.

Because this research constitutes a proof of concept based on a single case ($n = 1$), conclusions must be formulated cautiously. The value of this chapter lies not in advancing broadly generalisable claims, but in demonstrating what can be established in a methodologically defensible manner within the chosen research design. Chapter 6 thereby forms the bridge between the specific results and the broader discussion of the possibilities and limitations of AI-facilitated polyglot language acquisition.

§6.2 Answer to Research Question 1

Research Question 1: *To what extent can A2 level be achieved in three related languages within an accelerated, grammar-free learning programme with support from AI tools?*

Answer (placeholder): On the basis of the data presented in Chapter 5, it can be established that …

[The substantive answer will be inserted here once the empirical data are available.]

Interpretation. The results are interpreted in light of the design principles set out in Chapter 3. Consideration is given to the extent to which the structure of the learning programme — with its emphasis on input variation, scaffolding and consolidation — contributed to the development of functional language proficiency.

[Placeholder for interpretation: what do the observed trends imply for the feasibility of grammar-free learning and the role of AI as a supporting medium?]

Validity and reliability. The reliability of this conclusion rests on triangulation between different data sources (Anki, Speechling, logbooks and self-assessments). The interpretation must take account of:

- the *n = 1 design*, which means that the results cannot be generalised;

- possible *effects of motivation or consolidation factors;*
- the *reliance on self-report* within the autoethnographic context;
- and the *consistency of external assessors* in evaluating pronunciation and intelligibility.

Validity is strengthened by the internal consistency between the different measurement instruments, but remains provisionally limited to the context of this specific learning design.

§6.3 Answer to Research Question 2

Research Question 2: *How does active metalinguistic awareness contribute to recognising, distinguishing and correctly applying lexical and structural patterns within the learning programme?*

Answer (placeholder): On the basis of the data presented in Chapter 5, it can be concluded that …

[The substantive answer will be inserted here once the empirical results are available.]

Interpretation. The interpretation focuses on the role of metalinguistic awareness as a link between implicit learning and explicit pattern recognition. Consideration is given to the extent to which the design principles described in Chapter 3 — such as reflective prompts, interlingual comparison and structured feedback — promoted conscious analysis of similarities and differences between the three target languages.

[Placeholder for interpretation: what do the observed patterns imply for the functioning of the reflection formats, the use of AI feedback as a stimulus for thought, and the transfer from insight to application?]

Validity and reliability. The interpretation is based on triangulation between logbook reflections, annotations accompanying AI interactions and cloze exercises. In assessing validity and reliability, account must be taken of:

- the *self-reflective nature* of the data, which makes interpretation susceptible to subjective bias;
- possible *variation in depth* between weeks or languages;
- the *coding decisions* involved in the thematic analysis of reflections;
- and the *context dependence* — language task, prompt type and time pressure — under which metalinguistic insight became observable.

Reliability is strengthened through consistent coding and comparison across multiple sources of reflection data, but remains provisionally bounded by the individual case.

§6.4 Answer to Research Question 3

Research Question 3: *To what extent does structural relatedness between the target languages function as an accelerating factor in simultaneous acquisition, provided that it is explicitly exploited through interlingual comparison?*

Answer (placeholder). On the basis of the results presented in Chapter 5, it can be established that …

[The substantive answer will be inserted here once the empirical data are available.]

Interpretation. The interpretation focuses on how structural relatedness between Spanish, Italian and Portuguese influenced the learning process. It is examined whether interlingual comparison actually resulted in acceleration, error reduction and durable understanding of linguistic systematicity. The analysis is linked to the design principles set out in Chapter 3, in which relatedness is treated not merely as a linguistic property, but as an instructional mechanism.

[Placeholder for interpretation: what do the observed tendencies imply for the usefulness of relatedness as an instructional heuristic? Which patterns indicate positive or negative transfer, and what role did metalinguistic awareness play in this process?]

Validity and reliability. The reliability of the interpretation rests on comparison between production tasks, time measurements and logbook reflections. The assessment must take account of:

- possible **asymmetries between languages** (e.g. faster development in one target language);
- the **limited sample size** of controlled comparative tasks;
- **task difficulty and fatigue** as external influencing factors;
- and the **interpretative dependence** on reflective self-reporting.

Validity is supported by triangulation between quantitative and qualitative data, but generalisability remains limited to the specific design and the particular combination of languages examined.

§6.5 Answer to the Central Research Question

Together, the three empirical subquestions provide the basis for answering the central research question of this study:

"Under what conditions can AI-facilitated support contribute to accelerating the simultaneous acquisition of three related languages to CEFR level A2, without loss of grammatical understanding or functional proficiency?"

On the basis of the data analysed in Chapter 5, this question can be answered as follows:

[Placeholder: the definitive answer will be inserted here once all empirical results are available.]

The provisional interpretation suggests that the effectiveness of AI support depends not primarily on the volume of generated input, but on the way in which that input is embedded within a coherent learning design. When variation, repetition and reflection are appropriately balanced, simultaneous acquisition to A2 level appears feasible within an accelerated programme. A central condition is that AI systems do not operate autonomously, but remain subordinate to a human-designed structure in which variation in examples, spaced repetition and metalinguistic guidance reinforce one another.

Robustness of this conclusion. The strength of this answer lies in the internal consistency of the findings across the subquestions, but the robustness of this conclusion is limited by the nature of the research: an $n = 1$ design with a strong autoethnographic component. Validity and reliability have been discussed for the individual subquestions (§6.2–6.4); together, they determine the extent to which this provisional conclusion can be generalised. Further confirmation requires replication in larger cohorts and with other combinations of languages.

§6.6 Implications for Theory and Practice

The findings of this research have implications at three levels: theoretical, instructional and policy-related.

Theoretically, the research supports the proposition that simultaneous language acquisition does not necessarily require grammar instruction, provided that pattern recognition and reflection are explicitly supported. This creates a bridge between usage-based learning and contrastive analysis: language is acquired not only through exposure, but also through active comparison and differentiation.

Instructionally, the design provides a concrete proof of concept: a modular, reproducible learning programme in which AI tools (GPT, Anki, Speechling) function as part of the operational infrastructure. Their combined use demonstrates that input, feedback and reflection can be partly automated without relinquishing human direction.

At the policy level, the results suggest that compact, AI-supported learning programmes may offer a realistic alternative to traditional sequential language courses. A continuing condition is that the technology be embedded within a transparent framework that preserves learner ownership and transforms the role of the teacher from a source of knowledge into a designer of learning experiences.

Transition to Chapter 7. The answer to the central research question concludes the empirical phase. In the following chapter, the methodological and epistemological implications of using AI as an instructional medium and research instrument are discussed.

Box 6.x — Scenarios of Empirical Outcomes (Linked to the Evidential Requirements in Chapter 5)

Introduction. The scenarios below describe possible empirical outcomes of the learning programme in light of the evidential requirements established in Chapter 5: X_1–X_4 (functional proficiency / CEFR A2), Y_1–Y_4 (metalinguistic awareness / pattern recognition) and Z_1–Z_3 (relatedness as an accelerating factor). Conditions that are not met are treated as *red flags*. The scenarios function as an assessment framework for subsequent empirical interpretation.

Scenario A — Full convergence
X_1–X_4, Y_1–Y_4 and Z_1–Z_3 met.
All learning components operate in conjunction: A2 level achieved in three languages, conscious pattern recognition and targeted transfer.
Red flags: none.
Implication: the design is replicable; next step = scaling up and external validation.

Scenario B — Strong functional performance, limited reflection
X_1–X_4 met; Y_1–Y_2 partially met; Z mixed.
Functional proficiency is high, but reflections remain superficial.
Red flags: limited explicit articulation of rules; weak bridge between implicit ↔ explicit learning.
Implication: redesign reflection formats; add prompts for pattern comparison.

Scenario C — Strong reflection, slow functional development
Y_1–Y_4 and Z_1–Z_2 met; X_2–X_3 partially met.
Insight is strong, but automatisation is slow.
Red flags: low retention / high accuracy / low speed.
Implication: allow more time for consolidation and introduce faster feedback cycles.

Scenario D — Structural acceleration without metacognition
Z_1–Z_3 met; X_1–X_2 partially met; Y conditions not met.
Relatedness is exploited without corresponding insight.
Red flags: uncontrolled transfer; fossilisation of errors.
Implication: incorporate reflection points focused on transfer.

Scenario E — Lexical growth without structural understanding
X_2 partially met; remaining conditions not met.
Rapid vocabulary growth through Anki, but limited applicability.
Red flags: high cloze scores, low communicative performance.
Implication: introduce more contextual sentences and greater awareness of collocation.

Scenario F — Imbalance between languages
X_1–X_2 met in two languages but not in the third; Y_2–Y_3 partially met; Z incomplete.
One language becomes dominant in interlingual comparison.
Red flags: cross-language interference; preference for the dominant language.
Implication: revise the input ratio; strengthen the weaker language through contrastive work.

Scenario G — Overreliance on AI
Y_2–Y_3 partially met; X / Z unclear.
The learner relies too heavily on model feedback.
Red flags: low error awareness; lack of self-correction.
Implication: reduce AI interventions; introduce more autonomous tasks.

Scenario H — Reflective stagnation
Y_1–Y_2 met only intermittently; remaining conditions not met.
Motivation declines and reflections become repetitive.
Red flags: limited logbook input; low variation.
Implication: introduce new reflection prompts and additional variation.

Scenario I — Delayed convergence
X conditions partially met; Y and Z emerge only late in the programme.
Learning is positive but slow; relatedness is exploited only at a late stage.
Red flags: slow increase in cloze scores; delayed development of reflective depth.
Implication: extend exposure time / expand Week 0.

Scenario J — Design failure (non-functional programme)
No conditions met.
No learning gains; lack of coherence between instrumentation and objectives.
Red flags: negative retention curves; inconsistent AI output.
Implication: fundamental reconsideration of the principle of ‘grammar-free learning’.

Closing note
Scenarios A–J cover the full spectrum of possible empirical outcomes. In non-optimal scenarios, the *red flags* indicate which evidential requirements (X, Y or Z) have remained unmet. The framework enables rapid diagnosis during data analysis and provides a basis for targeted redesign.

This section \ box will be removed in the final dissertation.

Chapter 7 – Implications and Concluding Reflections

§7.1 From Design to Application

The preceding chapters have established the foundation for this research: Chapter 3 set out the learning design, Chapter 4 provided the methodological justification, and Chapter 5 presented a preliminary, model-based elaboration. Together, these elements constitute a coherent design that is ready for implementation in a real learning environment. The model-based elaboration served as a form of conceptual exploration: it illustrated how the assumptions developed in Chapters 2 and 3 translate into concrete learning pathways and possible outcomes, without making any claim to empirical validity.

The crucial next step is the empirical implementation of the learning book in practice, with data collection and analysis conducted in accordance with the strategies set out in Chapter 4. Only through such implementation can it be established to what extent the assumed conditions actually lead to accelerated and functional language acquisition. Section 7.1 therefore marks the transition from design and methodology to empirical testing: the question is no longer *whether* the design is feasible, but *under what conditions* it can be validated empirically.

§7.2 Implications for Further Research

The findings of this dissertation must be interpreted in light of the methodological choices made. These choices — autoethnography, an $n = 1$ design, a thirteen-week timeframe and a focus on three closely related Romance languages — have produced rich and in-depth data, but they also limit generalisability. These limitations, however, provide clear directions for further research.

First, the **autoethnographic approach** requires replication in broader contexts. Whereas this research centres on a unique combination of language learner and researcher, subsequent studies could apply the same design principles to groups of language learners or to different educational settings. This would make it possible to examine the extent to which the mechanisms described here — interlingual strategies, scaffolding and metalinguistic awareness — remain robust beyond a single case.

Second, the ***n* = 1 design** means that the findings are strongly embedded in individual reflection. Further research could incorporate triangulation by external assessors and independent evaluators, allowing intersubjective reliability to be established alongside internal consistency.

Third, the **validation of AI-generated material** in this research was conducted on a sample basis. Larger-scale studies could systematically examine the extent to which language models generate reliable and semantically parallel input, and under what conditions such material is suitable for use in language education.

Furthermore, the **twelve-week duration** places particular emphasis on short-term effects. Future research could investigate longer learning programmes in order to determine whether the effects of simultaneous language acquisition are also sustained over time.

Finally, the choice of **related languages (Spanish, Italian and Portuguese)** limits the scope of the conclusions. Further research could examine whether the principles identified here — particularly the role of transfer and interlingual comparison — also apply to less closely related languages, or even to languages with fundamentally different grammatical systems.

By pursuing further research along these lines, the scientific community can test the extent to which the design principles developed here remain valid in broader and more complex contexts. This dissertation therefore represents not an endpoint, but a point of departure for systematic research into AI-facilitated polyglot language acquisition.

In addition to these directions for further research, there are broader implications for theory and practice, which are discussed below.

§7.3 Implications for Theory and Practice

The findings of this dissertation extend beyond the specific case of the simultaneous acquisition of Spanish, Italian and Portuguese. Theoretically, they support the importance of interlingual comparison and metalinguistic awareness as potential accelerators of language acquisition. This creates a bridge between usage-based approaches and traditions of contrastive analysis: language is acquired not solely through exposure, but also through active reflection on similarities and differences.

For practice, the implications lie primarily in the domain of design principles. The combination of triplet-based learning materials, structured scaffolding and the integration of AI tools — such as Claude and ChatGPT, Anki and Speechling — provides a reproducible blueprint that can also be applied beyond this particular case. This demonstrates that AI need not function as a replacement for teachers, but can serve as a complement that facilitates variation, feedback and reflection.

Finally, there are also policy implications. Educational institutions seeking to offer polyglot learning programmes could focus on compact learning trajectories in which AI systems provide enriched input and stimulate reflection, provided that they are embedded within a carefully designed instructional framework.

§7.4 Reflection on the Scope and Limitations of the Design

The learning design developed in this dissertation is primarily aimed at achieving functional proficiency at A2 level in three related languages (Spanish, Italian and Portuguese). The instructional choices made — the use of semantically equivalent triplets, an emphasis on pattern recognition and systematic interlingual comparison — are closely aligned with the descriptors

at this level. The question arises, however, to what extent this design can also be scaled to higher levels (B1–C2).

An important consideration is that the requirements of CEFR levels **above A2** are fundamentally different in character. Whereas A2 focuses primarily on simple communicative situations and a limited vocabulary, higher levels require a much broader range of skills, including:

- **Register sensitivity:** the ability to adapt language use to context, audience and social situation. From B1 onwards, the language learner must be able to distinguish between formal and informal language use, while C1–C2 require subtle control of style and tone, for example in a job application letter as opposed to an informal message.
- **Discourse competence:** the ability to produce and understand longer, coherent texts. This extends beyond sentence level to include the structure of an argument, the use of cohesive devices such as linking expressions and references, and the development of a narrative or argumentative line. This becomes a core requirement from B2 onwards.

The current design — based on parallel examples, explicit strategies and inductive reflection — is particularly well suited to enabling beginning language learners to develop functional language proficiency rapidly. For A2 objectives, the instruments used (Anki, Speechling and GPT-facilitated comparison) are therefore appropriate and efficient. For higher levels, however, this is insufficient, because other dimensions of language competence become dominant: fluent discourse production, register variation and the ability to engage with abstract or specialised topics.

This reflection makes clear that the principal strength of the design lies in accelerating the initial phase of language acquisition. The challenge for further research is to explore whether — and how — the principles of pattern comparison and strategy-oriented learning can be integrated into learning environments aimed at B1–C2. This will probably require additional forms of learning activity, such as task-based communication in extended interactions, discourse-level writing tasks and register-specific scenarios.

§7.5 LLMs and the Possibility of an Interlingual Grammar

A recurring problem in polyglot language acquisition is that grammars exist for individual languages, but no systematic description exists for an entire language family. Attempts to develop family-wide grammars or intercomprehension projects, such as EuroComRom, remain fragmentary: they provide strategic entry points or selective patterns, but not an exhaustive system that treats the Romance languages as a single object of learning. A theoretical foundation on which a simultaneous learning programme could rest is therefore still lacking.

The emergence of large-scale language models such as LLMs raises the question of whether this gap might eventually be filled. GPT models are not rule-based grammars, but statistical pattern-recognition systems capable of exposing emergent relationships on the basis of very large corpora. In principle, this makes them suitable for identifying structural correspondences and divergent developments across several languages simultaneously. Whereas the polyglot language learner develops an emergent network of patterns within a mental interlingual system, an LLM could externalise and systematise such patterns.

A hypothetical research programme could therefore involve deliberately supplying LLMs with parallel corpora of related languages and using them to identify generalisations that have so far existed only implicitly. This could produce the first contours of an **interlingual grammar**: a metagrammar tied not to a single language, but to the language family as a whole. This has not yet been achieved. LLMs in their current form are pre-trained predictors that produce fragments of interlingual observation rather than a comprehensive grammatical system.

Future research could, however, examine whether fine-tuning or retrieval-based augmentation makes it possible to construct such family-wide grammars. AI could thereby become not only an instructional tool, but also a new instrument of knowledge production within comparative linguistics. For polyglot learning, this would introduce a possible third option alongside the opposition between 'grammar-free' and 'rule-based' learning: learning from an AI-generated emergent rule space derived not from the literature, but directly from the statistical relationships within a language family.

§7.6 Concluding Reflections

This dissertation set out to explore how AI-supported polyglot language acquisition to A2 level can be designed and investigated. The central aim was not to present a ready-made method, but to develop a research design that creates the conditions for systematic evaluation. Chapter 2 provided the theoretical foundation, Chapter 3 translated this into a learning design, and Chapter 4 described the methodological instrumentation. Chapters 5 and 6 provided the framework for subsequent empirical completion.

Chapter 7 has provided the reflection needed to place these elements in perspective. The contribution of this dissertation lies not in the volume of data or the breadth of its application, but in the attempt to explore a new paradigm: multilingual learning in which AI is incorporated as a functional component, with the learning process remaining central and technology serving as a means rather than an end.

The strength of this work lies in its integrated approach: theory, design and methodology have been interwoven to produce a framework that is testable, reproducible and extensible. Its limitations are equally apparent: the scale is small, the empirical basis remains provisional, and reliance on generative AI introduces uncertainties that can only be addressed systematically through further research.

Nevertheless, the combination of insights, design choices and methodological rigour supports the proposition that this dissertation contributes to rethinking language acquisition in an era in which AI is becoming increasingly prominent. It is not merely a description of an experiment, but an invitation to dialogue: how does human creativity relate to artificial generation, and what new forms of learning become conceivable as a result?

If AI provides any insight into language acquisition, it is this: acceleration appears possible, provided that it is embedded within a design that deliberately creates space for reflection and structuring. This insight is consistent with the usage-based tradition and with research on metalinguistic awareness and scaffolding; the evidence currently available suggests plausibility, although the proposition remains open to empirical challenge.

Epilogue – Duifs and Canarian

At the end of a dissertation, there are supposed to be acknowledgements. But words such as "love" and "support" sound so flat in academic language — as though they could have been written by a system. Yet I know this work would not have existed without *her*, without my great love, my Dove. Not without *Hare Duivigheid* — 'Her Doveliness' — who reminded me every day that language only comes alive when someone hears it.

Together we speak a language that is registered nowhere: *Duifs*, the language of the *Duiven* - in our case, the language of 'Love Doves'. It is neither a Spanish dialect nor a secret code, but simply a variety of Dutch - with a sound and rhythm that only we recognise. A language that once began as a joke and grew into a way of life. Anyone who hears it thinks of birds, and rightly so: *Duifs* shares the air with Canarian. One language is spoken by an island full of people, the other by just two. Fifteen years ago, on Tenerife, I casually asked a passer-by: *¿Hablar Canario?* That question, half joke, half wonder, became the beginning of a line of inquiry that has never let me go: how animal language once developed into human language — and how meaning took wing.

Duifs is our language of nuance, more a matter of *pluisibiliteit* - 'pluisibility' in Anglo-Dutch - than of *plausibility*. In pigeon logic, something need not be plausible as long as it feels *pluis*: a Dutch word suggesting that things are safe, sound and as they should be. The test is not reasoning, but atmosphere: is everything *pluis*? That is how we speak to one another in trust, and it has proved to be a surprisingly robust model of communication.

DUMOR - *Duiven Humor*, or 'Pigeon Humour' - is her speciality: a mixture of lightness and logic that makes life bearable. Anyone unfamiliar with it recognises it immediately: "I applied for a job at the animal welfare society, but I turned it down — it sounded like a real dog of a job." That is DUMOR: innocent wordplay with a soft landing. It is humour that does not bite, but comforts; language that flies alongside you for a moment before coming in to land.

Hare Duivigheid is the alpha and omega of this dissertation: the reason why language here is not only its subject, but also its destination. Without her, I might perhaps have learnt *about* language, but I would not have learnt to speak *in* language again. And that, I now know, is the difference between knowledge and understanding.

And so I end where it began: with wonder at the capacity of language to connect worlds. Pigeons have been doing so for centuries with their wings. We did it with words.

And that is why I dedicate this work to my great love: to Hare Duivigheid — for whom I learnt to speak Duifs.

Remco Schimmel,

Summer 2026

References

1. Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.
2. Bentz, C., Alikaniotis, D., Cysouw, M., & Ferrer-i-Cancho, R. (2017). The impact of morphological complexity on learning. *Journal of Language Modelling, 5*(2).
3. Bentz, C., & Berdicevskis, A. (2016). Language complexity across domains. *Linguistic Typology, 20*(2).
4. Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.
5. Bickel, B., & Nichols, J. (2013). *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology.
6. Bickerton, D. (1990). *Language and species*. University of Chicago Press.
7. Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2).
8. Bonami, O., & Boyé, G. (2022). **[Full bibliographic record still to be verified.]**
9. Bono, M., & Varela, [initials still to be verified]. (2021). **[Full bibliographic record still to be verified.]**
10. Bonvino, E. (2009). Cognates and lexical processing in Romance languages. *Journal of Romance Studies, 9*(2).
11. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*.
12. Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
13. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1).
14. Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. Diller (Ed.), *Individual differences and universals in language learning*. Newbury House.
15. Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. Psychological Corporation.
16. Cenoz, J. (2000). Research on multilingual acquisition. In J. Cenoz & U. Jessner (Eds.), *English in Europe: The acquisition of a third language*. Multilingual Matters.
17. Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching, 46*(1).
18. Cenoz, J., & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *The Modern Language Journal, 95*(3).
19. Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. John Benjamins.
20. Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
21. Colpaert, J. (2020). Reflections on present-day CALL research. *ReCALL, 32*(2).
22. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
23. De Bot, K., & Jaensch, C. (2015). What is special about L3 processing? *Bilingualism: Language and Cognition, 18*(2).

24. De Bot, K., & Lowie, W. (2020). **[Full bibliographic record still to be verified.]**
25. Deacon, T. (1997). *The symbolic species: The co-evolution of language and the brain*. Norton.
26. de Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*(1).
27. Dewaele, J.-M., & Wei, L. (2013). Is multilingualism linked to a higher tolerance of ambiguity? *Bilingualism: Language and Cognition, 16*(1).
28. Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology.
29. Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2).
30. Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Blackwell.
31. Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics—Introduction to the special issue. *Applied Linguistics, 27*(4).
32. Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3).
33. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10).
34. Futrell, R., Mahowald, K., & Gibson, E. (2015). Dependency locality as a predictor of processing difficulty. *Proceedings of the National Academy of Sciences, 112*(33).
35. Galambos, S. J., & Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition, 34*(1).
36. García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
37. Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1).
38. Godwin-Jones, R. (2018). Using mobile technology to develop language skills and cultural understanding. *Language Learning & Technology, 22*(3).
39. Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press.
40. Gombert, J. E. (1992). *Metalinguistic development*. University of Chicago Press.
41. Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.
42. Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598).
43. Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics*. Multilingual Matters.
44. Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.
45. Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh University Press.
46. Jessner, U. (2008). Teaching third languages: Findings, trends and challenges. *Language Teaching, 41*(1).
47. Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed. draft).
48. Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*(5).

49. Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*.
50. Kusters, W. (2003). *Linguistic complexity: The influence of social structure on language structure*. LOT.
51. Lee, J. H. (2021). Artificial intelligence in language education: A critical review. *Computer Assisted Language Learning, 34*(5–6).
52. Lee, J. S., & Drajati, N. A. (2019). A review of technology-mediated L2 pragmatic development. *System, 83*.
53. Li, S. (2015). The construct validity of language aptitude. *Studies in Second Language Acquisition, 37*(1).
54. Li, Z., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*.
55. Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.
56. Lim, Y., Choi, Y., & Lee, H. (2023). Human–AI collaboration in education: Challenges and opportunities. *Computers & Education, 200*, 104792.
57. Liu et al. (2023). **[Full bibliographic record still to be verified.]**
58. Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition*. Academic Press.
59. Lynch, T. (2001). Seeing what they meant: Transcribing as a route to noticing. *ELT Journal, 55*(2).
60. Maiden, M. (2011). *The Romance Languages*. Cambridge University Press.
61. Maiden, M. (2016). *The Romance Languages*. Oxford University Press.
62. Maiden, M., & Robustelli, C. (2007). *A reference grammar of modern Italian*. Routledge.
63. Mayer, T., & Cysouw, M. (2012). Creating a comparative database of cross-linguistic lexical distances. In *Proceedings of the Workshop on Linked Data in Linguistics*.
64. Meara, P., & Rogers, V. (2019). *Vocabulary and the study of language aptitude*. Routledge.
65. Mehisto, P. (2012). *Excellence in bilingual education: A guide for school principals*. Cambridge University Press.
66. Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5).
67. Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
68. Odlin, T. (2003). Cross-linguistic influence. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Blackwell.
69. Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. Routledge.
70. Otwinowska, A. (2016). *Cognate vocabulary in language acquisition and use: Attitudes, awareness, activation*. Multilingual Matters.
71. Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4).
72. Posner, R. (1996). *The Romance languages*. Cambridge University Press.
73. Reynolds, B. L. (2023). Generative AI in second language education: Affordances and challenges. *Language Learning & Technology, 27*(2).
74. Ringbom, H. (2007). *Cross-linguistic influence in second language acquisition*. Multilingual Matters.

75. Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.
76. Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism, 13*(2).
77. Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model of L3 acquisition. *Bilingualism: Language and Cognition, 18*(2).
78. Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294).
79. Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4).
80. Serva, M., & Petroni, F. (2008). Indo-European languages tree by Levenshtein distance. *Europhysics Letters, 81*(6), 68005.
81. Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
82. Skehan, P. (2016). Foreign language aptitude and its relationship with grammar. *Language Teaching, 49*(3).
83. Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.
84. Steiner, G. (2011). *The poetry of thought: From Hellenism to Celan*. New Directions.
85. Swain, M. (1985). Communicative competence: Some roles of comprehensible input and output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition*. Newbury House.
86. Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching, 48*(1).
87. Tenney, I., Das, D., & Pavlick, E. (2019). BERT rediscovers the classical NLP pipeline. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.
88. Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
89. Tomasello, M. (2008). *Origins of human communication*. MIT Press.
90. Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard University Press.
91. VanPatten, B. (2015). *While we're on the topic: BVP on language, acquisition, and classroom practice*. ACTFL.
92. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
93. Wang, Y., & Vasquez, C. (2012). The effect of written corrective feedback by L1 and L2 speakers on ESL learners' development. *System, 40*(2).
94. Weber, M. (2002). *Science as a vocation* (H. H. Gerth & C. W. Mills, Trans.). Hackett. (Original work published 1919)
95. Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2).
96. Yilmaz, Y., & Granena, G. (2019). Task-based oral corrective feedback in computer-mediated communication. *Language Learning & Technology, 23*(2).
97. Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2).
98. Zhai, X. (2022). The role of paraphrasing in second language acquisition: An AI-supported study. *Computer Assisted Language Learning, 35*(3).

Glossary — AI-facilitated Polyglot Language Acquisition

Preferred term	Variants / related terms	Definition / context of use
AI-facilitated support	AI support	The use of generative AI as part of the learning process, including for the provision of language input, comparison, feedback and temporary support. AI does not constitute an autonomous instructional system, but functions within a human-designed learning architecture.
AI-driven material production	AI-generated learning materials	The production of learning and practice materials with generative AI according to pre-specified substantive, semantic and formal requirements. Generated material is not used as learning or research input without prior control.
Cumulative divergence	accumulation of divergence	The occurrence of more than one semantic or structural point of divergence within a single exercise unit. Cumulative divergence is avoided because it prevents unambiguous identification of which difference the language learner is perceiving or processing.
Controlled variation	designed variation	The deliberate modification of one pre-specified linguistic or constructional dimension while keeping other relevant characteristics as constant as possible. This allows a pattern to become visible without multiple differences simultaneously interfering with the comparison.
Grammar-free acquisition	grammar-free learning	Language acquisition without explicit instruction in grammatical rules. Grammatical knowledge is assumed to be capable of developing gradually through exposure, pattern recognition, comparison, repetition and reflection. *Grammar-free* therefore does not mean that grammatical structure is absent, but that it is not primarily taught as an explicit system of rules.
Intercomprehension	receptive multilingualism	The ability to understand related languages without fully mastering each of them separately, by systematically drawing on shared vocabulary, formal similarities and structural relatedness.
Interlingual strategies	strategy codes / comparison techniques	Systematically identified patterns used to recognise similarities and differences between Spanish, Italian and Portuguese. At word level, the strategies distinguished are M (morphological), F (phonetic), G (graphic), S (systematic suffixes), T (transfer), D (divergence) and W (alternating consonants). Sentence-structural patterns are coded separately as Z-strategies (Z1–Zx).
Interlingual grammar	family-wide metagrammar	A theoretical concept for a grammatical description that does not centre on a single language, but on systematic correspondences and divergences within a group of related languages. The research explores whether LLMs can help externalise and organise such patterns from parallel language data.
Latin as an interlingual key	Latin key / etymological anchor	The use of Latin as a genealogical and heuristic reference framework for structural relationships between Romance languages. Latin is not an independent learning target. In material production, a Latin anchor is used only when the derivation is correct, transparent and directly defensible.

Preferred term	Variants / related terms	Definition / context of use
Metalinguistic awareness	metalinguistic reflection	The ability consciously to perceive and identify patterns, similarities and differences between languages and to make use of them when processing new language input. In this research, it functions both as a hypothesised learning mechanism and as a variable to be investigated empirically.
Micro-context	exclusive context	A short sentence context designed to be semantically specific enough for the target lemma to constitute the only plausible completion. If the lemma is replaced by a semantically close alternative, the sentence must become incorrect or clearly implausible.
Design grammar	production rules / design rules	The set of explicit rules according to which practice material is constructed. These include sentence length, semantic parallelism, lemma exclusivity, constraints on variation, strategy coding and rules governing divergence. The design grammar determines the properties that individual exercise units must satisfy.
Production architecture	production and control system	The broader organisation of AI-driven material production in which design, generation, derivation, control, correction and verification are functionally distinguished. The production architecture includes, among other things, small production batches, bounded contexts, serial derivation and separate verification steps.
Reflection formats	reflection questions / reflection tasks	Standardised tasks through which the language learner records observed patterns, differences, transfer and learning experiences. They serve both an instructional function — stimulating metalinguistic awareness — and a research function by generating qualitative data.
Scaffolding	temporary support / gradual guidance	The temporary provision of support that is progressively reduced as the language learner becomes increasingly able to recognise and apply patterns independently. Within the learning programme, this support may consist of controlled contexts, repetition, hints, comparative material, AI feedback and reflection tasks.
Semantically equivalent triplet	triplet / parallel sentences	A set of three parallel sentences in Spanish, Italian and Portuguese expressing the same event, state or conceptual content. Meaning must remain fully parallel across the three languages; differences should serve only to make the linguistic form or strategy under investigation visible.
Serial derivation	derivation from an anchor language	A production method in which a single reference or anchor sentence is designed first and the corresponding sentences in the other languages are derived directly from it. This avoids unconstrained parallel generation and reduces the risk of semantic drift between the languages.
Week 0	receptive initial phase / preparatory week	The intensive preparatory phase at the beginning of the thirteen-week learning programme. In Week 0, 600 lemmas in each of three languages — 1,800 language forms in total — are presented receptively to activate lexical, phonological and morphological pattern recognition. The primary objective is orientation and pattern recognition, not full active mastery of the lexicon.
Self-recording learning design	self-documenting research design	A research design in which learning activities simultaneously generate data about the learning process. Exercises, performance, errors, retention and reflections therefore function both as components of the intervention and as sources for empirical analysis.