Cognitive and Multimodal Perspectives in Contemporary English Philology: Theory, Analysis, and Applications

Khodorenko, A.

Dnipro

Serednyak T.K.

2025

Khodorenko, A. *Cognitive and multimodal perspectives in contemporary English philology: Theory, analysis, and applications*. Dnipro: Serednyak T.K. – 73 p. ISBN 9-786178540579

Cognitive and Multimodal Perspectives in Contemporary English Philology: Theory, Analysis, and Applications
This monographic edition explores the intersection of cognitive linguistics and multimodal discourse within the framework of contemporary English philology. It provides theoretical foundations, analytical tools, and practical applications for studying language, communication, and meaning making across multiple semiotic modes. The work is aimed at scholars and advanced students interested in cognitive approaches and multimodal analysis in linguistic research.

Acknowledgments

I am sincerely grateful to my closest family, friends, and teachers, both near and far, whose unwavering support and encouragement have been invaluable throughout this journey. Their inspiration has continually reminded me of the importance of perseverance, reflection, and intellectual curiosity. They have given me strength during difficult moments, encouraged me to think critically, and inspired me to explore new ideas with an open heart and mind.

I would also like to extend my deepest appreciation to those who have shared their insights and wisdom with me over the years, helping to shape both my personal and academic growth. Without their belief in me, this work would not have been possible.

THE TABLE OF CONTENTS

Preface

This book, *Cognitive and Multimodal Perspectives in Contemporary English Philology: Theory, Analysis, and Applications*, emerges from an intellectual journey deeply rooted in the humanistic tradition. It was inspired by a longstanding desire to explore how we make sense of language—not only through words, but through images, gestures, spatial design, and the many multimodal forms of meaning-making that shape our everyday and academic lives.

The precursors to this work are many: from the foundations laid by cognitive linguists such as Lakoff, Johnson, and Langacker, to the semiotic insights of Kress, van Leeuwen, and Halliday. But even more broadly, the motivation stems from a humanistic impulse—to understand language as a deeply human act, shaped by experience, culture, and creativity. In this sense, the book is not merely an academic study, but an invitation to think, interpret, and reflect through a lens that values complexity, context, and interpretation.

Humanistic inquiry has always emphasized reading between the lines. In a world increasingly dominated by algorithms, visual overload, and compressed forms of communication, the ability to think critically and empathetically has never been more essential. This book advocates for that kind of reflective reading and writing, grounded in philological traditions but extending into new multimodal terrains.

Art plays a key role in this paradigm shift. From the expressive gestures captured in Rubens' paintings to the evolving narratives in digital visual media, this work seeks to rethink how we engage with the aesthetic and rhetorical functions of communication. Rubens, with his dramatic compositions and human emotions rendered in oil, reminds us that meaning is never flat or linear—it is layered, embodied, and deeply contextual. So too must our analytical approaches become attuned to the multimodal richness of meaning in contemporary texts.

As readers move through these chapters, I hope they are not only informed but also inspired to approach texts—whether linguistic, visual, digital, or performative—with a renewed sense of curiosity, humanity, and critical awareness. For in doing so, we keep alive a vital humanistic tradition: the art of interpretation.

TARGET AUDIENCE AND RELEVANCE OF THE WORK

This volume is designed to address the needs of a diverse yet interconnected academic and professional audience engaged in the study and application of language in both theoretical and practical contexts. The content and structure of the book reflect its interdisciplinary scope, making it especially valuable to researchers in applied and cognitive linguistics, English language pedagogy, translation studies, and discourse analysis.

Foremost among the intended readers are researchers in applied and cognitive linguistics, for whom the book offers both theoretical insights and empirical applications. As the fields of linguistics continue to evolve toward more integrated and cognitively grounded models, this work contributes to ongoing discussions about the role of conceptual metaphor, multimodal representation, and cognitive structures in language processing and use. The inclusion of methodological reflections and case studies enhances its relevance for scholars involved in both foundational research and data-driven analysis.

Another key audience comprises teachers of English as a Foreign Language (EFL) and curriculum developers, particularly those interested in incorporating recent findings from cognitive science into their pedagogical practices. By presenting cognitive grammar, metaphor-based instruction, and multimodal analysis as practical tools for classroom implementation, the book bridges the gap between linguistic theory and language teaching methodology. This supports the growing interest in cognitively informed didactics, especially in contexts where learners must navigate abstract, idiomatic, or specialized language.

The book is equally suited for graduate students in philology, linguistics, education, and related disciplines who are developing their academic competence in language theory, discourse analysis, and translation. The clarity of exposition, coupled with critical discussions and applied examples, makes it accessible to readers seeking to deepen their understanding of how cognitive and multimodal theories are reshaping traditional approaches to English philology and language education.

Finally, translators and discourse analysts will find particular value in chapters dedicated to specialized terminology, metaphor in political discourse, and the role of multimodal elements in meaning construction. These sections offer theoretical and methodological guidance for analysing the interplay between language, context, and cognition in translation practices and critical discourse work.

By addressing the needs of these varied audiences, the book promotes interdisciplinary dialogue and advances the practical utility of theoretical

knowledge, demonstrating how insights from cognitive linguistics and multimodality can inform and transform multiple domains within the study and use of language.

PART I

INTRODUCTION

The theoretical underpinnings of linguistic and semiotic analysis have undergone significant evolution in recent decades, shifting away from purely formal and structural paradigms toward more cognitively and contextually grounded frameworks. Among these, Cognitive Linguistics (CL) has emerged as a key approach that views language as inherently linked to human cognitive processes, such as perception, categorization, and conceptualization (Langacker, 1987; Lakoff & Johnson, 1980). Rather than treating language as an autonomous system, CL emphasizes its embodiment, experiential basis, and its role in reflecting and shaping thought.

Parallel to this cognitive turn in linguistics, the field of Multimodality has developed as an interdisciplinary framework that explores how meaning is constructed not solely through verbal language, but through the interplay of diverse semiotic resources, including visual, auditory, spatial, and gestural modes (Kress & van Leeuwen, 2001). Multimodality complements and extends the insights of CL by acknowledging that human communication is fundamentally multisensory and contextual, with meaning emerging from the dynamic integration of various modes in situated discourse.

Together, these theoretical perspectives offer a robust foundation for the study of terminology and conceptual structures**,** particularly in applied domains such as scientific communication, translation, and educational linguistics. Terminological units are not isolated lexical items but are embedded in larger cognitive and semiotic networks that must be interpreted across linguistic and non-linguistic modalities. The combination of CL and multimodal theory allows for a nuanced exploration of how abstract and technical concepts are mentally represented, linguistically encoded, and visually or spatially articulated.

In this theoretical context, the present work explores the intersection of Cognitive Linguistics and Multimodality as it applies to the analysis and creation of terminological glossaries**,** with a special focus on environmental science and water protection. By examining how conceptual knowledge is structured, represented, and communicated across multiple modes, this investigation contributes to the growing body of research that bridges cognitive science, language studies, and semiotics. The aim is not only to deepen theoretical understanding but also to support practical outcomes in the design of more effective educational and communicative tools.

Another foundational principle of Cognitive Linguistics is **embodiment**, the notion that our bodily experiences shape the way we conceptualize the world and, consequently, the way we use language. According to this view, linguistic meaning arises not only from abstract mental processes but also from sensorimotor experiences. For instance, spatial concepts such as *up*, *down*, *in*, and *out* often extend metaphorically into domains such as emotion and social status—e.g., *feeling down* or *climbing the social ladder*. These metaphorical extensions are not arbitrary but reflect deeply ingrained patterns of thought grounded in physical and cultural experience (Lakoff & Johnson, 1980). This perspective reinforces the idea that meaning is not merely referential but is constructed through interaction with the environment and shaped by cultural norms.

CL also emphasizes the **usage-based** nature of language. Rather than viewing linguistic competence as a static knowledge of rules, CL considers linguistic knowledge to be emergent from language use. Frequent exposure to particular patterns leads to entrenchment—where constructions become cognitively entrenched over time and form the basis for grammatical knowledge (Langacker, 1987; Bybee, 2010). This perspective aligns CL with corpus linguistics and other empirical approaches that rely on real-world language data to understand how linguistic structures develop, evolve, and vary across contexts and speakers.

In terms of application, CL has been influential in several fields, including discourse analysis, cognitive poetics, semantics, and translation studies. In translation, for example, cognitive linguistic tools such as conceptual metaphor theory help to explain how metaphors are structured differently across languages and cultures, revealing the cognitive constraints on cross-linguistic equivalence (Kövecses, 2005). In the domain of literary studies, cognitive poetics uses CL frameworks to explore how readers mentally simulate narratives, interpret figurative language, and construct meaning from literary texts.

Furthermore, CL has contributed significantly to the understanding of terminology in specialized domains. Since terms are closely linked to conceptual structures, cognitive approaches to terminology study how expert knowledge is organized and lexicalized within a domain. This is particularly relevant in fields like environmental science or medicine, where neologisms and term formation are closely tied to conceptual innovation. Here, CL provides tools for analysing how technical concepts are framed, categorized, and metaphorically structured, enhancing both terminological research and communication strategies across disciplines.

Cognitive Linguistics offers a holistic and interdisciplinary approach to language that integrates meaning, usage, and cognition. By emphasizing embodiment, conceptualization, and the dynamic nature of linguistic knowledge, CL provides a powerful alternative to formalist theories and opens new avenues for linguistic inquiry and application across both theoretical and applied domains.

METHODOLOGY AND METHODOLOGICAL FRAMEWORK

The methodology adopted in this book is rooted in a qualitative, interdisciplinary framework that brings together cognitive linguistics, multimodal theory, English philology, and applied linguistics. Rather than focusing on quantification or statistical generalization, this approach privileges depth, interpretation, and contextual understanding. It seeks to explore the nuanced ways in which meaning is constructed, negotiated, and communicated in English through the interaction of various semiotic modes—linguistic, visual, auditory, and spatial.

Qualitative methods are particularly well-suited to the study of multimodal communication because they allow for rich, contextualized interpretations of how language operates in situated discourse. This book uses discourse analysis, multimodal analysis, content analysis, and semiotic interpretation to investigate selected texts and communicative events. These methods facilitate close readings of texts—ranging from political speeches and media narratives to literary works and digital genres—highlighting the ways in which verbal and non-verbal resources work in tandem to shape mental representations, ideologies, and social practices.

From a philological perspective, this methodological orientation enables a historically and culturally sensitive reading of English texts, both contemporary and historical. It allows for the tracing of how patterns of metaphor, framing, and multimodal expression have evolved across time and how they reflect shifts in worldview, power structures, and cultural cognition. Philological sensitivity also supports the exploration of intertextuality, stylistic variation, and genre conventions in relation to multimodal meaning-making.

The book adopts a case study approach, analysing selected examples in depth rather than seeking large-scale representativeness. Each case is chosen to exemplify a particular communicative phenomenon or to illustrate how multimodal elements function within a given context. These case studies are informed by theoretical frameworks from cognitive semantics (e.g., conceptual metaphor theory), social semiotics, and systemic functional linguistics, ensuring that interpretation is grounded in robust scholarly traditions.

Furthermore, the methodology includes visual and multimodal annotation—a qualitative process of coding that accounts for gestures, images, spatial layout, sound, and typography. This enables a layered analysis of how meaning is not only said but shown, felt, and enacted, particularly in English as it functions in globalized, mediated, and technologically embedded contexts.

In sum, the qualitative orientation of this research prioritizes meaning over measurement, complexity over simplification, and interpretation over enumeration. It reflects a commitment to understanding English not merely as a linguistic code, but as a multimodal, culturally embedded, and cognitively mediated system of representation.

RESEARCH METHODS

1. **Corpus Analysis**

One of the primary tools used in this book is corpus analysis. Corpora, both contemporary and historical, will be analyzed to identify patterns in the use of multimodal elements, including metaphorical language, visual markers, and spatial arrangements. This analysis will focus on both written and spoken English texts, from a wide range of genres including news reports, advertisements, political speeches, literary works, and academic articles.

The corpus will be coded for instances where verbal and non-verbal modes (e.g., images, layout, prosody) work together to construct meaning. Specialized tools for corpus annotation, such as AntConc and other linguistic software, will allow for the identification and categorization of specific linguistic patterns (e.g., metaphors, collocations) in conjunction with their multimodal counterparts (e.g., images, spatial organization).

2. **Discourse Analysis**

Discourse analysis will be employed to examine how multimodal elements shape the way meaning is constructed in both oral and written communication. This method will look at how various forms of communication (linguistic, visual, gestural, etc.) collaborate to produce meaning and convey power, ideology, and identity. In this sense, discourse analysis will provide a lens to understand how multimodal communication contributes to the construction of social realities.

This analysis will draw on frameworks such as Critical Discourse Analysis (CDA), which focuses on understanding how power and ideologies are reflected and perpetuated in discourse. This is particularly important in examining how multimodal communication in English can reflect hegemonic ideologies, particularly in media, politics, and

advertising. CDA will help identify the underlying values encoded in multimodal texts, such as those concerning race, gender, class, and nationalism.

3. **Multimodal Analysis**

The core methodology for analysing how different semiotic modes function together will be multimodal analysis, as proposed by Kress and van Leeuwen (2001). This approach will allow for the systematic examination of how different modes (language, images, gestures, spatial organization) work together to convey meaning. Multimodal analysis involves breaking down the elements of a text into its constituent semiotic modes and analysing how these elements interrelate to construct meaning.

In the case of multimodal texts such as advertisements, political speeches, and media coverage, the focus will be on how visual elements (e.g., color, framing, imagery), spatial elements (e.g., layout, positioning), and auditory elements (e.g., tone, emphasis, rhythm) work alongside linguistic features to reinforce or challenge the verbal message. By integrating multimodal analysis with cognitive linguistics, the book will examine how metaphors, in particular, are not only linguistically but also visually and spatially represented.

4. **Case Studies**

Case studies will be utilized throughout the book to apply the theoretical framework and methods to real-world examples. These case studies will involve in-depth analyses of multimodal texts from various domains, such as political discourse, advertising, news media, and literature. For example, the analysis of political speeches may examine how metaphors like “the war on terror” or “the battle against poverty” are reinforced by visual imagery (e.g., flags, maps, images of soldiers) and how these multimodal elements work together to shape public opinion.

In literature, a case study might explore how 19th-century novels use spatial organization and visual imagery in conjunction with language to communicate themes of power and social status. Similarly, advertisements will be analysed to see how product imagery, layout, and linguistic elements collaborate to construct consumer identities and reinforce cultural norms.

For the analysis outlined in the methodology, a wide range of corpora will be collected and analysed, focusing on multimodal communication in English. These corpora will serve as the primary data sources for the investigation of linguistic, visual, and other semiotic modes. The first category of corpora includes linguistic resources. The British National Corpus (BNC) is a comprehensive resource of both written and spoken

English from the late 20th century, providing a balanced representation of various registers such as spoken language, fiction, academic texts, and newspapers. This corpus will be used to investigate the prevalence of metaphors, linguistic structures, and their interaction with other semiotic modes. The Corpus of Contemporary American English (COCA) contains texts from a wide range of genres, including fiction, newspapers, academic articles, and spoken language. It will be employed to examine multimodal communication patterns, including metaphorical language, across different contexts in American English. The International Corpus of English (ICE) offers a unique resource for comparing English usage in various countries, which will help highlight cultural variations in multimodal communication and metaphor use in diverse English-speaking communities. This corpus is especially valuable for understanding how English and its multimodal practices manifest in non-native contexts.

The second category is multimodal corpora, which includes resources that combine text with visual elements. The Multimodal Corpus of Political Discourse includes political speeches, debates, and media coverage that blend text with visual elements like charts, images, and graphs. This corpus will be analyzed for its use of metaphors and the co-existence of non-verbal cues, such as gestures, tone of voice, and facial expressions, in political communication. It will help uncover how ideologies are conveyed through the combination of written and visual elements in political discourse. The Multimodal Composition and Communication Corpus (M3C) contains multimedia texts, such as advertisements, promotional materials, and websites, which incorporate various modes like text, images, sound, and video. This corpus will be analyzed to explore how advertising campaigns and online platforms use multimodal communication to convey messages, influence behavior, and create cultural meanings. News Media Corpora include texts from newspapers, online articles, and news broadcasts that combine written language with visual elements such as photos and videos. These corpora will be particularly useful for studying how news outlets use multimodal framing to shape public opinion on topics like politics, health crises, and social issues. For instance, media coverage of COVID-19 often combined textual reports with visual representations of the virus, medical professionals, and public reactions.

Finally, the third category includes digital and online corpora. The Twitter and Social Media Corpus provides rich multimodal data in the form of short textual posts combined with images, videos, and hyperlinks from platforms like Twitter, Instagram, and Facebook. These corpora offer insight into how social media users combine multiple modes of communication to create and share meaning in online interactions.

THEORETICAL FRAMEWORK

The theoretical framework underpinning this book is drawn primarily from cognitive linguistics (CL), multimodal communication theory, and English philology. Cognitive Linguistics, with its focus on conceptual structures, embodied cognition, and metaphorical mappings, provides the foundation for understanding how humans conceptualize the world through language. In particular, Conceptual Metaphor Theory (CMT) will play a central role in explaining how abstract concepts are structured and conveyed through metaphorical language, supported by multimodal resources.

Simultaneously, multimodality theory, particularly as articulated by Kress and van Leeuwen (2001), guides the analysis of non-verbal semiotic modes (images, gestures, spatial arrangements, typography) and how these modes interact with language to produce meaning. In this view, meaning is not solely a product of linguistic elements, but emerges through the interplay between multiple modes that co-function in communicative contexts. This perspective is crucial for understanding how language operates not just as a system of words but as part of a broader semiotic system where different modes contribute to the full expression of meaning.

Finally, English philology provides the necessary lens for examining historical and contemporary texts in English. The focus on philology in this book is crucial for understanding how language has evolved over time and how multimodal communication has been embedded within English texts across different periods, genres, and contexts. The integration of philological principles with cognitive and multimodal theories provides a rich methodological approach for studying the evolution of English and the various ways it has interacted with visual and non-verbal elements throughout its history.

A critical aspect of CL is the idea that linguistic patterns are shaped by human experience and usage. According to CL, language is not a static system of abstract rules but is dynamic and evolves based on the cognitive experiences of individuals. This view of language as grounded in experience makes CL highly relevant for applied fields such as language teaching, where the goal is to connect linguistic forms with the lived experiences and conceptual knowledge of learners. Unlike formalist theories that emphasize abstract rules and structures, CL focuses on how language is shaped by the real-world experiences of speakers, emphasizing the role of context and usage in shaping linguistic forms. As a result, CL highlights the importance of context in meaning construction and challenges the notion that meaning can be fully explained through rigid formal systems.

One of the most influential contributions of CL is the theory of image schemas, which are fundamental cognitive structures that underlie much of human understanding. Image schemas are basic, recurring patterns of experience that shape how humans conceptualize the world. These schemas are not learned consciously but are instead deeply ingrained in cognition. They form the building blocks for more complex thought processes and are reflected in the language we use. For example, spatial image schemas like Container**,** Path, and Source-Goal provide the basis for understanding spatial relationships and are often reflected in the metaphors and grammatical structures of a language. The Container schema, for instance, underlies expressions such as "She's in the room," "He put the book in the box," and "I'm feeling trapped." These expressions all rely on the conceptualization of space as an enclosed area, a conceptual structure that is deeply rooted in human cognition. The study of these schemas allows linguists to better understand how people conceptualize the world and how language reflects these conceptualizations.

Another central concept in CL is conceptual metaphor theory (CMT), developed by George Lakoff and Mark Johnson. According to Conceptual Metaphor Theory (CMT), metaphors are not merely decorative elements of language but serve as fundamental mechanisms of human thought and cognition. Lakoff and Johnson (1980) argue that metaphors structure our understanding of abstract concepts by systematically mapping them onto more concrete, embodied experiences. For example, the metaphor *time is money* conceptualizes the abstract notion of time in terms of the more tangible domain of economics, as seen in expressions such as "I'm running out of time," "She spends her time wisely," or "He invested a lot of time in that project."

Beyond this widely cited example, a range of other conceptual metaphors illustrates how deeply metaphor shapes cognition across various domains. In emotional discourse, the metaphor *anger is heat* is reflected in expressions such as "She exploded with rage," "He was boiling with anger," or "That comment ignited his fury." In the domain of epistemology, *understanding is seeing* is a pervasive metaphor, evident in phrases like "I see your point," "That's a clear idea," or "Let me shed some light on that." Social status is often understood through the metaphor *status is verticality*, which underlies expressions like "She's at the top of her field," "He fell from grace," or "They rose quickly through the ranks."

In communicative and argumentative contexts, the metaphor *ideas are physical objects* emerges in statements like "Let's shape that proposal," "They threw out the suggestion," or "We need to construct a solid argument." Relationships are frequently conceptualized through the metaphor *relationships are journeys*, seen in expressions such as "We're going through a rough patch," "They're at a crossroads," or "Our

relationship has gone off track." Similarly, life itself is often understood via the metaphor *life is a journey*, as in "He's at a turning point," "She's on the right path," or "They've come a long way."

These examples demonstrate that metaphor is not confined to literary or rhetorical usage, but is instead a pervasive and indispensable tool in human conceptualization. By analysing such metaphorical mappings, Cognitive Linguistics sheds light on the cognitive structures that underlie language use and abstract reasoning, contributing to our understanding of how meaning is constructed across cultures and contexts.

The emphasis on meaning and conceptual structure in CL has significant implications for applied linguistics, particularly in the field of language teaching. By emphasizing the role of experience, context, and conceptual structures in shaping language, CL offers a framework for understanding how language learners acquire and use language. Traditional language teaching methods often focus on grammar rules and vocabulary without considering the underlying conceptual structures that shape meaning. CL, on the other hand, encourages educators to focus on how learners conceptualize the world and how this influences their use of language. By integrating insights from CL into language teaching, educators can help learners build a deeper understanding of the relationship between language and thought.

Moreover, CL's emphasis on usage-based models challenges the notion that language learning is purely a matter of memorizing rules and forms. Instead, CL emphasizes the importance of exposure to real language use in context. This approach aligns with contemporary pedagogical theories that stress the importance of authentic language use and communicative competence. In this context, language learners are encouraged to engage with language in meaningful ways, connecting linguistic forms with their own experiences and conceptual knowledge. This shift from a rules-based approach to a meaning-based approach has far-reaching implications for language teaching and for the ways in which linguistic competence is understood and assessed.

Cognitive Linguistics offers a paradigm shift in the way we understand language, its structure, and its relationship to cognition. By emphasizing the connection between language and general cognitive processes, CL provides a more holistic understanding of linguistic phenomena, focusing on meaning, experience, and conceptualization. Its insights into the role of metaphor, categorization, and image schemas offer valuable tools for understanding how language reflects and shapes human thought. Moreover, the application of CL principles in language teaching provides a framework for developing more effective, experience-based pedagogies that can help learners make deeper connections between language and thought.

CONCEPTUAL METAPHOR THEORY: THE FRAMEWORK OF LAKOFF AND JOHNSON

Multimodality really opens up a whole new angle on how we understand conceptual metaphor. For a long time, we looked at metaphor mainly through language — words on a page, phrases in a sentence. But when we try to grasp how meaning is actually built in the mind, that purely verbal perspective ends up being pretty limited. Sure, language can point us toward conceptual structures, but on its own, it often feels like just the surface layer. The deeper, richer layers — the embodied, emotional, and perceptual aspects — tend to get lost if we don't consider nonverbal modes too.

That's where multimodal approaches come in and change the game. When we think of meaning as something that unfolds across multiple channels — visual, gestural, spatial, even sonic — we suddenly have much more to work with. A gesture, an image, a sound — these can all carry metaphorical weight and help structure how we conceptualize complex or abstract ideas. Think of how a simple hand movement, a shift in tone, or a layout on a slide can "say" something that language alone can't. They don't just decorate meaning — they shape it.

Before, trying to scientifically visualize what happens in the mind when we form conceptual representations — especially metaphors — was a bit of a nightmare. But multimodality gives us tools to approach that. Instead of guessing, we now have ways to *construct* and even *reconstruct* those mental gestalts of meaning. We can track how different modes of expression interact to produce layered significance — lexical, syntactic, symbolic, even literary.

So multimodality doesn't just add colour — it adds depth**.** It shows how metaphors live in embodied experience, in shared cultural codes, and in the rich tapestry of modes we use to express ourselves. A metaphor in a novel isn't just in the words; it's in the rhythm, the imagery, the structure of the scene. Likewise, in a scientific diagram or political cartoon, metaphor can be embedded in layout, contrast, or movement — all nonverbal carriers of conceptual meaning.

This opens up incredible potential for analysis. We can start asking not just *what* a metaphor means, but *how* it's being built across modes — how sound, image, gesture, and text all collaborate to produce that spark of insight or resonance. And that takes us closer to how people actually experience metaphor in real life — not just as linguistic artifacts, but as felt, seen, enacted forms of understanding.

In this way, multimodal metaphor theory gives us not only a broader lens but also more precise tools — for linguistics, cognitive science, semiotics, and literary analysis alike. It invites us to reimagine meaning as something dynamic, embodied, and distributed — and to study

metaphor not as a dead metaphor on a page, but as a living, moving, multisensory thing.

This reconceptualization of metaphor as dynamic and multisensory aligns closely with the foundations of Conceptual Metaphor Theory (CMT), a pivotal framework within Cognitive Linguistics that similarly views metaphor as central to human thought and understanding.

A cornerstone of Cognitive Linguistics is Conceptual Metaphor Theory (CMT), originally developed by George Lakoff and Mark Johnson in their seminal work *Metaphors We Live By* (1980). CMT challenges the traditional view of metaphor as a purely rhetorical or ornamental feature of language and instead presents it as a core mechanism of human cognition. According to this theory, metaphors allow individuals to understand abstract or unfamiliar concepts by mapping them onto more concrete, bodily-grounded experiences. Through this process of conceptual mapping, metaphor becomes essential to meaning-making, enabling individuals to make sense of complex or intangible domains.

While CMT was initially formulated within a linguistic framework, its principles have increasingly found resonance within multimodal research. The idea that conceptual metaphors structure thought suggests that metaphorical cognition is not confined to verbal language alone but can be expressed and reinforced through visual, gestural, spatial, and auditory modes. For instance, the metaphor "ARGUMENT IS WAR" is not only reflected in expressions such as "He shot down my argument" or "Her claims were indefensible," but also frequently visualized in political cartoons, news graphics, or advertisements where debate is portrayed as combat, complete with imagery of weapons, battlefields, or opposing forces. Similarly, the metaphor "TIME IS MONEY" often appears in digital interfaces, where countdown clocks, transactional metaphors, and financial icons visually encode the urgency or value of time.

This convergence between CMT and multimodality highlights the embodied and experiential nature of conceptual metaphors. Since many source domains draw upon sensory and spatial experiences—such as motion, balance, containment, or orientation—these metaphors are naturally suited to cross-modal representation. Visual metaphors in advertising, layout strategies in digital texts, or the alignment of gestures in political speeches are all examples of how metaphor extends beyond verbal articulation and into multimodal expression. A metaphor such as "LIFE IS A JOURNEY," for instance, may be linguistically realized through phrases like "He's at a crossroads," but also visually represented through paths, roads, or vehicles in film, literature, and branding.

The multimodal extension of CMT reinforces the theory's relevance to philological and linguistic inquiry, particularly in contexts where texts

are no longer purely verbal. Contemporary philology must account for how meaning is co-constructed through various semiotic resources, and conceptual metaphors offer a unifying cognitive mechanism through which these modes interact. Moreover, the consistent mappings across modes suggest that metaphor serves not only as a linguistic strategy but also as a design principle in meaning-making processes. In educational contexts, awareness of both conceptual metaphors and their multimodal realizations can enhance students' ability to interpret complex texts, visual narratives, and persuasive discourse.

In sum, integrating CMT with multimodality broadens the explanatory power of metaphor as a cognitive and communicative tool. It affirms that metaphor is not merely embedded in language, but in the coordinated deployment of multiple semiotic systems. This intersection invites new research in applied linguistics, translation studies, discourse analysis, and digital philology, where metaphors shape not only how we speak and write, but also how we visualize, gesture, and design communication in increasingly multimodal environments.

Metaphorical Framing in Contemporary Discourse

In contemporary media and political discourse, metaphorical language serves as a cognitive and communicative tool to simplify complex phenomena. During the COVID-19 pandemic, war metaphors became a dominant narrative device, with expressions such as "frontline workers," "battling the virus," and "fighting the pandemic" casting the health crisis in terms of militaristic engagement. While this framing sought to emphasize urgency and collective action, it has been critiqued for potentially oversimplifying the multifaceted nature of public health and contributing to heightened societal anxiety.

Similarly, sports metaphors frequently appear in political journalism, with terms such as "political football," "game-changer," and "level playing field" rendering complex power dynamics in familiar, competitive terms. Though such language can enhance accessibility, it may inadvertently trivialize substantive policy debates.

In economic discourse, metaphors such as "overheating economy" or "fiscal cliff" help readers grasp abstract or technical market behavior through embodied or spatial schemas. These expressions translate economic instability into tangible imagery, guiding interpretation and emotional response.

Environmental communication often relies on metaphors to convey scientific data to a general audience. Phrases like "carbon footprint," "greenhouse effect," and "tipping point" are commonplace. While useful in illustrating abstract environmental processes, these metaphors can also mislead if interpreted literally or without adequate context.

Finally, crime reporting employs metaphorical frames that can significantly influence public sentiment and policy. Referring to crime as a "virus" invokes notions of contagion and containment, often aligning with preventive or rehabilitative policy preferences. In contrast, framing crime as a "beast" suggests inherent aggression and may lead to more punitive societal responses.

Table 1. Metaphorical Domains and Functions in Public Discourse

Domain	Common Metaphors	Source Domain	Communicative Purpose	Potential Impact
Public Health (COVID-19)	"Frontline workers," "battle the virus," "fighting the pandemic"	War	To mobilize action and emphasize urgency	May induce anxiety, militarize public discourse
Politics	"Political football," "game-changer," "level playing field"	Sports	To simplify competition and strategy in governance	Risk of trivializing serious political issues
Economics	"Overheating economy," "fiscal cliff"	Physical States, Spatial Motion	To visualize economic processes and risks	Can exaggerate threats, shape public economic expectations
Climate Change	"Carbon footprint," "greenhouse effect," "tipping point"	Environment, Balance	To conceptualize scientific phenomena in relatable terms	May distort scientific accuracy without clarification
Crime Reporting	"Crime is a virus," "crime is a beast"	Disease, Animal Behavior	To frame causes and responses to criminal behavior	Shapes public opinion toward prevention or punishment

MULTIMODALITY IN LINGUISTIC AND PHILOLOGICAL RESEARCH

The growing recognition of language as a complex, embodied, and context-sensitive phenomenon has led to a significant expansion in how meaning is understood and studied in the humanities. One of the most influential paradigms to emerge in recent decades is that of multimodality, which views communication as a process that extends beyond verbal language to include a wide array of semiotic resources such as gesture, image, sound, spatial arrangement, and typography. In the context of linguistic and philological research, the integration of

multimodal theory offers a powerful lens through which to revisit long-standing questions and to generate new methodologies for the analysis of texts—both historical and contemporary.

Multimodality challenges the long-standing assumption that language is the sole or primary carrier of meaning. Instead, it posits that communicative acts are typically realized through the coordinated interaction of multiple modes, each contributing in distinctive ways to the construction and interpretation of meaning. For example, in a spoken conversation, gesture, prosody, facial expression, and spatial positioning may play as crucial a role in meaning-making as the words themselves. In written texts, layout, font size, punctuation, and accompanying visuals can significantly influence how the reader interprets the message. This recognition prompts a shift in philological and linguistic inquiry—from focusing exclusively on verbal language to examining how meaning emerges through a synergy of semiotic resources.

Incorporating multimodal analysis into linguistic and philological research offers several critical benefits. First, it allows for a more nuanced understanding of historical texts. Medieval manuscripts, illuminated texts, and early printed works often relied on a rich interplay of visual and textual elements, from marginalia and iconography to color-coded rubrics and stylized scripts. A multimodal approach makes it possible to analyze how these non-verbal features contribute to the semantic, rhetorical, and interpretive dimensions of the text. By attending to these elements, researchers can uncover layers of meaning that might otherwise be overlooked in a purely textual analysis.

Second, multimodal theory proves indispensable in the analysis of contemporary media, where the integration of multiple modes is the norm rather than the exception. In digital communication—ranging from websites and video essays to social media posts and infographics—meaning is constructed through the simultaneous orchestration of visual, auditory, textual, and sometimes tactile modes. In this context, traditional philological approaches require adaptation in order to remain analytically relevant. Multimodality provides the theoretical and methodological tools needed to address this complexity, equipping researchers with frameworks for dissecting how modes interact to shape interpretation.

The implications of multimodality also extend to the study of specialized discourse, including scientific, political, and educational texts. In technical manuals, for instance, diagrams and layout are not supplementary but integral to the instructional content. In political discourse, visual metaphors and performative gestures often reinforce or subvert spoken messages. In classroom settings, multimodal pedagogy encourages learners to engage with content through various sensory channels, fostering deeper cognitive processing and retention. Multimodal analysis, therefore, enhances the study of discourse by

accounting for the full range of communicative resources that shape understanding in different contexts.

Moreover, multimodality opens new avenues for empirical research in applied linguistics and translation studies. For translators, an awareness of multimodal dimensions is crucial when working with audiovisual material, advertising, or graphic novels—genres where linguistic choices are closely tied to visual and auditory cues. Similarly, corpus linguistics and digital humanities projects increasingly incorporate tools for analysing multimodal data, allowing scholars to trace patterns of multimodal usage across large datasets. This intersection of multimodality with digital technologies reinforces the paradigm's relevance in current and future research.

Finally, the theoretical convergence between multimodality and cognitive linguistics provides fertile ground for interdisciplinary exploration. Both frameworks emphasize the embodied, experiential, and situated nature of meaning-making, and both challenge the notion of language as an abstract, disembodied system. While cognitive linguistics focuses on conceptual structures such as image schemas, metaphors, and frames, multimodality extends the analysis to the semiotic instantiation of these structures across different media and modes. This synergy offers a comprehensive account of how individuals construct, interpret, and negotiate meaning in real-world communicative practices.

In conclusion, multimodality represents a paradigm shift in how language and meaning are understood within linguistics and philology. By moving beyond the confines of verbal text, it provides scholars with a richer, more holistic framework for analysing the interplay of modes in communication. As media environments continue to evolve, and as academic disciplines increasingly embrace interdisciplinarity, multimodal approaches will play a crucial role in shaping the future of linguistic and philological inquiry.

DEFINING MULTIMODALITY

Multimodality in Language and Communication

Multimodality refers to the simultaneous and integrated use of multiple semiotic modes in the production and interpretation of meaning. This paradigm has emerged as a response to the limitations of monomodal theories that focus predominantly on verbal language, often neglecting the ways in which communication occurs through a combination of visual, gestural, spatial, and aural modes (Kress & van Leeuwen, 2001; Norris, 2004). From this perspective, meaning is not exclusively conveyed through words but is the result of a coordinated interplay among various semiotic resources.

Kress and van Leeuwen (2001) famously proposed that "all communication is multimodal," arguing that even written texts rely on layout, typography, and visual structuring to convey meaning. Their social semiotic approach extends Halliday's (1978) systemic functional linguistics, applying the metafunctional model—ideational, interpersonal, and textual—to multiple modes beyond language. In this model, visual and other non-linguistic resources are not secondary or supplementary but co-equal systems of meaning. Similarly, Bateman (2008) emphasized that different semiotic modes contribute specific affordances, and their integration often follows culturally and contextually motivated patterns.

Multimodal discourse analysis (MDA), as developed by scholars such as Scollon and Scollon (2003) and Jewitt (2009), operationalizes this theory by offering tools to examine how communicative meaning emerges from the orchestration of modes. For instance, gestures, gaze, body positioning, color schemes, images, and sound are all analytically significant, particularly when examining complex media such as websites, films, educational materials, or political speeches. MDA thus reveals meaning as a composite construction, wherein each mode contributes different kinds of information that would be incomplete or ambiguous if interpreted in isolation (Jewitt, 2014).

This understanding directly challenges the traditional, logocentric assumptions in philology and linguistics, which historically have privileged written and spoken text as the primary conveyors of meaning (Bezemer & Kress, 2008). In classical philological analysis, visual and material aspects of manuscripts—such as illuminations, marginalia, or script—were often treated as peripheral. Multimodality, however, repositions these elements as integral to textual meaning, leading to more holistic interpretations of historical and contemporary artifacts (Machin, 2016).

Multimodality also aligns with broader cognitive and embodied views of language. Researchers such as Forceville (2006) and Fauconnier and Turner (2002) have demonstrated that metaphors and mental spaces are often instantiated across multiple modes, reinforcing the claim that cognition itself is inherently multimodal. This has profound implications for applied linguistics, translation studies, and education, where multimodal literacy—understanding and producing meaning across modes—is increasingly recognized as an essential competence (Kalantzis & Cope, 2012).

In sum, multimodality represents a paradigm shift in how language, communication, and meaning are understood. By recognizing the integrated and interdependent roles of different semiotic modes, it not only expands the scope of linguistic and philological inquiry but also provides new methodologies for engaging with the complexity of

contemporary and historical discourse. As communication becomes increasingly mediated by digital technologies and visual culture, the relevance of multimodal frameworks is likely to grow, offering valuable insights into both the form and function of human meaning-making.

Multimodality in English Language and Communication: Toward a Consistent Mental Representation of the World

The notion that language merely reflects reality has long been challenged by developments in cognitive linguistics, semiotics, and discourse studies. Rather than passively encoding information, language—particularly as used in English, a global lingua franca—acts as an active constructor of mental models of the world. Multimodality extends this principle by recognizing that meaning is not solely encoded in verbal structures but emerges through the interaction of various semiotic modes, such as visual layout, gesture, sound, space, and typography. These modes co-construct a coherent mental representation of experience that is cognitively processed and socially situated.

Language as One Mode Among Many

Historically, English philology and linguistics have focused primarily on written and spoken language as the central modes through which meaning is encoded and interpreted. This textual and oral bias reflects long-standing traditions in both classical philology and structuralist linguistics, which emphasized grammatical structures, lexical semantics, and phonological systems as the primary loci of analysis. However, with the advent of what has been termed the "multimodal turn" in discourse and communication studies (Kress & van Leeuwen, 2001; Jewitt, 2009), language is increasingly conceptualized not as an isolated code, but as one semiotic mode among many that jointly contribute to the production of meaning.

In contemporary communicative practices, especially in English, meaning is rarely conveyed through verbal language alone. In both formal and informal contexts, communication is embedded in multimodal environments in which elements such as gesture, gaze, image, sound, spatial arrangement, and typography interact with linguistic signs to shape interpretation. For instance, a political speech does not derive its persuasive power solely from its lexical choices or syntactic constructions. Rather, its impact depends on an integrated performance involving prosodic modulation, facial expressions, physical posture, visual symbolism (e.g., flags, uniforms, backdrops), and spatial organization (such as the speaker's placement in relation to the audience or media). These non-verbal resources contribute not only to rhetorical

effect but also to the activation of shared cultural scripts and emotional frames (Charteris-Black, 2011).

Similarly, in educational settings, classroom discourse often combines spoken language with diagrams, gestures, body orientation, and the use of digital tools like interactive whiteboards or presentation software. Such multimodal scaffolding supports the construction of complex ideas and facilitates knowledge transmission by engaging multiple channels of cognitive processing (Bezemer & Kress, 2010). This is especially evident in disciplines such as science or mathematics, where abstract concepts often require visual metaphors, spatial representations, and embodied demonstrations for effective understanding.

The growing emphasis on multimodality aligns closely with embodied and distributed models of cognition. According to theories proposed by Clark and Chalmers (1998) and further developed by scholars such as Fauconnier and Turner (2002), cognition is not confined to the internal processes of the brain but is distributed across the body, artifacts, and social environments. This theoretical framework posits that humans use language, gesture, visual aids, and physical objects as cognitive extensions that enable them to represent and manipulate complex mental content.

From this perspective, multimodality provides the necessary representational flexibility to account for how English speakers and writers construct shared cognitive spaces—or what Johnson-Laird (1983) called "mental models"—for abstract domains such as time, causality, emotion, morality, or ideology. For example, metaphors such as "argument is war" or "time is money" are not merely linguistic expressions but are grounded in embodied experiences and extended through multimodal resources, including images, spatial diagrams, and narrative structures (Lakoff & Johnson, 1980; Forceville, 2008).

In sum, multimodal analysis broadens the scope of philological inquiry and linguistic interpretation by acknowledging that meaning in English is fundamentally co-constructed across verbal and non-verbal modes. This shift demands methodological and theoretical frameworks capable of addressing the full complexity of communicative acts as they occur in real-world, situated contexts.

Consistency Through Conceptual Integration

A central function of multimodal communication is to preserve internal coherence in how individuals construct and share representations of the world. In any communicative act, multiple semiotic resources—verbal, visual, gestural, spatial, and auditory—contribute simultaneously to meaning-making. The integration of these resources into a unified interpretive structure relies on consistent conceptual correspondences across modes. This process is supported by semiotic relations,

particularly metaphoric, iconic, and indexical links, which allow for cross-modal mappings that maintain internal consistency within a mental model or scene.

The theoretical basis for this integration is provided by Conceptual Metaphor Theory (CMT) (Lakoff & Johnson, 1980; Kövecses, 2010) and Conceptual Blending Theory (CBT) (Fauconnier & Turner, 2002). These frameworks explain how abstract domains are understood through mappings from more concrete source domains, often in ways that transcend the verbal mode. For example, the conceptual metaphor TIME IS A JOURNEY structures temporal understanding in spatial terms. This metaphor can manifest linguistically through expressions such as "We are approaching the deadline" or "We've fallen behind schedule." However, its multimodal realization is equally pervasive: timeline graphics, progress bars, motion arrows, and countdown clocks all reinforce the spatial structure of temporal experience. These visual cues are not redundant but rather essential in anchoring the verbal metaphor within a perceptual frame that enhances comprehension and retention.

In English-language media and communication—particularly in domains such as advertising, science communication, and educational discourse—such multimodal metaphors have become highly conventionalized. For instance, infographics about climate change frequently combine spatial metaphors (e.g., rising sea levels depicted by upward movement), visual intensity markers such as color gradients (ranging from green to red to signal increasing danger), and linguistic hedging strategies ("could lead to severe consequences") to construct a composite mental space. This integration of modes facilitates the communication of complex or unfamiliar content by linking it to embodied, familiar schemas rooted in everyday experience (Forceville, 2008; Bateman, 2014).

Furthermore, these cross-modal mappings do more than illustrate or embellish linguistic content; they shape the conceptual structure of what is communicated. As such, they are not supplementary but constitutive of meaning itself. The ability to draw on multiple semiotic resources allows communicators in English to manage abstraction, navigate ambiguity, and frame ideologically charged content with greater rhetorical precision. Multimodal coherence thus becomes a cognitive affordance—one that reflects how human beings naturally think in integrated, non-linear, and perceptually grounded ways (Johnson, 2007; Kress, 2010).

References

Bateman, J. A. (2008). *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents.* Palgrave Macmillan.

Bateman, J. A. (2014). *Text and Image: A Critical Introduction to the Visual/Verbal Divide.* Routledge.

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166–195.

Bezemer, J., & Kress, G. (2010). *Multimodality, Learning and Communication: A Social Semiotic Frame.* Routledge.

Bybee, J. (2010). *Language, Usage and Cognition.* Cambridge University Press.

Cap, P. (2013). *Proximization: The Pragmatics of Political Discourse.* John Benjamins.

Charteris-Black, J. (2011). *Politicians and Rhetoric: The Persuasive Power of Metaphor* (2nd ed.). Palgrave Macmillan.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Danesi, M. (2016). *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet.* Bloomsbury.

Fauconnier, G., & Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* Basic Books.

Forceville, C. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In G. Kristiansen, M. Achard, R. Dirven, & F. J. Ruiz de Mendoza Ibáñez (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives* (pp. 379–402). Mouton de Gruyter.

Forceville, C. (2008). Metaphor in pictures and multimodal representations. In R. W. Gibbs Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 462–482). Cambridge University Press.

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning.* Edward Arnold.

Jewitt, C. (2009). *The Routledge Handbook of Multimodal Analysis.* Routledge.

Jewitt, C. (2014). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis* (2nd ed., pp. 15–30). Routledge.

Johnson, M. (2007). *The Meaning of the Body: Aesthetics of Human Understanding.* University of Chicago Press.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Harvard University Press.

Kövecses, Z. (2005). *Metaphor in Culture: Universality and Variation.* Cambridge University Press.

Kövecses, Z. (2010). *Metaphor: A Practical Introduction* (2nd ed.). Oxford University Press.

Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication.* Routledge.

Kress, G., & van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication.* Arnold.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By.* University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites.* Stanford University Press.

Machin, D. (2016). *Introduction to Multimodal Analysis* (2nd ed.). Bloomsbury.

Mayer, R. E. (2005). *The Cambridge Handbook of Multimedia Learning.* Cambridge University Press.

Norris, S. (2004). *Analysing Multimodal Interaction: A Methodological Framework.* Routledge.

Scollon, R., & Scollon, S. W. (2003). *Discourses in Place: Language in the Material World.* Routledge.

PART II

MULTIMODALITY IN DIGITAL ENGLISH

In the 21st century, the global dominance of English in digital spaces has fundamentally reshaped how communication unfolds, particularly through the lens of multimodality. Digital environments—ranging from social media platforms and online journalism to educational content and professional communication—are inherently multimodal. This means that meaning is no longer carried by linguistic forms alone, but emerges through the orchestration of multiple semiotic modes such as visual imagery, spatial organization, gesture (in video), audio, and even interactivity. In such environments, the traditional linguistic message is supported, enhanced, or even subverted by other modes, creating layers of meaning that are best understood through a multimodal framework (Kress, 2010; Adami, 2015).

The ubiquity of multimodal texts in digital English reflects broader shifts in literacy practices and communicative norms. As users increasingly access information via screens, reading is no longer a linear, alphabetic activity but a non-linear, multisensory experience shaped by layout, color, motion, sound, and navigation pathways. For instance, a social media post is not just a short textual utterance—it may include emojis, hashtags, hyperlinks, images, and embedded videos. Each of these elements carries meaning potential and functions within specific genre and platform conventions. Emojis, for example, can signal tone, emotional nuance, or even irony, often disambiguating textual content that might otherwise be misinterpreted in the absence of prosodic or facial cues (Danesi, 2016; Page, 2018).

Multimodal communication in digital English is not merely decorative; it serves essential cognitive and rhetorical functions. For example, in political discourse on platforms such as Twitter, users frequently engage in complex acts of stance-taking and framing. A single tweet critiquing a policy may combine sarcastic phrasing, visual memes, quotation marks, and intertextual references (e.g., hashtags like #fail or #spin), each contributing to a collective, recognizable rhetorical position. These multimodal cues activate shared cultural knowledge and mental models, facilitating rapid interpretation among users who belong to the same interpretive community (Zappavigna, 2012).

Educational contexts further illustrate the cognitive benefits of multimodality in digital English. Online learning platforms often present content through combinations of simplified written text, voice-over narration, animations, diagrams, and interactive exercises. These multimodal configurations are not random but carefully designed to align with how people learn. According to Mayer's (2005) cognitive theory of

multimedia learning, presenting information through complementary channels (verbal and visual) reduces cognitive load, increases retention, and fosters deeper conceptual understanding. For instance, a video explaining climate change might combine narration with moving charts, symbolic colors (e.g., red for heat), and embedded text labels. Such integration supports learners in constructing coherent mental models of abstract scientific phenomena, particularly when these phenomena are difficult to visualize using language alone.

Moreover, digital multimodality enables global accessibility. English-language content on platforms such as YouTube, Duolingo, or Coursera often caters to non-native speakers. To ensure comprehension across diverse audiences, creators adopt a range of multimodal strategies: slower speech rates, subtitles, pictorial scaffolding, repetition, and gesture. These strategies compensate for linguistic complexity and promote inclusivity, allowing more users to engage with content that would otherwise remain inaccessible. Multimodality, in this context, becomes a pedagogical tool and a democratizing force in global English communication (Bezemer & Kress, 2008).

At a structural level, digital multimodal texts follow distinct genre conventions that shape user expectations and guide interpretation. For example, in Instagram stories or TikTok videos, storytelling often unfolds through a hybrid of spoken narration, text captions, background music, and on-screen gestures. These multimodal elements are synchronized to create affective engagement and narrative coherence. The temporal and spatial orchestration of semiotic modes in such texts reveals the increasing complexity and richness of English usage in digital environments, where communicative competence now requires multimodal literacy alongside traditional linguistic skills (Jewitt, 2009; Serafini, 2014).

Ultimately, understanding English in the digital age necessitates a multimodal lens. Language, while central, functions as one element within a broader semiotic ensemble. Whether users are posting memes, designing infographics, or navigating instructional videos, meaning is co-constructed across modes. As such, any analysis of digital English that privileges linguistic content alone risks missing the full communicative picture. The study of multimodality offers tools to explore this dynamic landscape, highlighting how English is not just written or spoken—but enacted, performed, and experienced through complex intermodal interactions.

Multimodality, Identity, and Ideology

The construction of mental representations is a dynamic process shaped by a complex interplay of language, visual imagery, and other semiotic modes. It is never a neutral act, as it is always mediated by ideological,

cultural, and power-laden factors. In the context of English language communication, multimodal communication plays a central role in shaping how individuals and groups perceive the world and themselves. Through the deliberate use of visual framing, linguistic register, spatial organization, and metaphor, meaning is not simply conveyed but constructed in ways that reflect and reproduce specific worldviews, ideologies, and power relations.

Consider the phrase "illegal alien." On its surface, the term might appear to be a simple descriptor for individuals who have entered a country without legal authorization. However, when placed within a broader multimodal context, the term takes on more complex ideological dimensions. The word "illegal" carries with it connotations of criminality and deviance, while "alien" invokes metaphors of invasion and otherness, positioning the subject as an outsider to be feared or controlled. In many media representations, this framing is further amplified by visual elements such as images of fences, border patrols, and confrontational scenes between migrants and authorities. These images do not simply accompany the text but co-construct the meaning, reinforcing and intensifying the ideological implications of the words. The effect is a composite representation that does more than describe a situation—it shapes how the audience interprets the situation and forms opinions about the individuals involved.

In this way, multimodal analysis becomes a powerful tool for critical discourse studies, as it allows researchers to examine how ideologies are encoded not just in verbal language, but in the ways in which language is coupled with visual and spatial elements. The combination of words, images, and spatial arrangement can produce new, often subtle, layers of meaning that guide the audience's emotional and intellectual responses. For instance, the visual placement of an image in a news article—whether it is prominently displayed above the fold or relegated to a small corner—can influence how seriously the issue is taken or how sympathetic the viewer is toward the subject. These multimodal choices reveal the ideological underpinnings of communication, showing how power dynamics are perpetuated through communication practices that are rarely questioned by the audience.

Moreover, as English has become a global language, it carries with it hegemonic modes of representation that extend beyond the linguistic choices made by individual speakers or writers. The use of standardized templates in news graphics, business presentations, and international diplomacy is one such example. These templates not only serve as tools for transmitting information but also implicitly shape how people should think, act, and feel. For instance, in international diplomacy, the framing of a leader's speech through visual cues—such as the background design, the clothing choices of the participants, or even the arrangement of the

furniture in the room—may convey a message of authority, control, or subservience that is not explicitly stated in the words themselves.

These hegemonic modes of representation are powerful because they shape the mental models through which individuals understand the world. Mental models are cognitive frameworks that help people make sense of new information, and they are heavily influenced by repeated exposure to certain communicative practices. Multimodal communication, through its repeated use of certain visual and linguistic conventions, stabilizes these mental models and embeds them into cultural cognition. For example, the consistent use of specific metaphors in media coverage of political leaders, such as framing a leader as "a beacon of hope" or "a strongman," shapes public perception of their character and leadership style. Over time, these metaphors become normalized and taken for granted, influencing how people understand political authority and power.

The influence of multimodal communication on mental models is especially important in the context of globalization. English as the global lingua franca carries with it not only linguistic forms but also cultural assumptions and ideological structures that are often masked as neutral or universal. The standardized modes of representation found in global media, advertising, and corporate communication are deeply rooted in Western-centric worldviews and practices. By perpetuating these conventions, multimodal communication serves to reinforce the dominance of certain ideologies—such as capitalism, individualism, and consumerism—across diverse cultural contexts. As such, critical multimodal analysis becomes a crucial tool for understanding how power and ideology are mediated through global communication practices.

In summary, multimodal communication is far more than the mere presentation of information through multiple modes—it is a central mechanism through which ideologies are constructed, maintained, and transmitted. The ways in which language, images, and spatial elements are combined and framed are not neutral; they actively shape how individuals perceive themselves, others, and the world around them. In the case of English as a global language, these multimodal practices carry with them hegemonic representations that implicitly shape global cultural norms, mental models, and power structures. By analysing multimodal communication critically, scholars and practitioners can gain insight into the often hidden ways in which language and other semiotic resources shape our collective understanding and actions.

Table 2. Illustrative table of the influence of multimodal communication in constructing mental representations and ideologies.

Example	Verbal Language	Visual Elements	Interpretation & Ideology	Effect on Public Perception
1. "Illegal Alien"	"Illegal Alien"	Images of fences, border patrol, confrontation	The term "illegal" connotes criminality, and "alien" emphasizes otherness; visual elements amplify the fear and threat associated with immigration	Creates an image of migrants as criminals or invaders, shaping negative public attitudes toward immigration
2. Political Discourse - "The Strongman"	"Strongman Leader"	Images of leaders standing assertively, surrounded by flags or military paraphernalia	The phrase emphasizes power and control, while visual cues (military imagery, flags) reinforce authority	Positions the leader as a figure of unquestionable authority and control, influencing public support or fear
3. "War on Drugs"	"War on Drugs"	Images of soldiers, police raids, drug paraphernalia	The metaphor frames drug addiction as a battle or war, while images of violence and conflict suggest urgency	Shapes the public view of drug addiction as an enemy to be fought, influencing policies towards more punitive measures
4. "Economic Recovery"	"Economic Recovery"	Graphs showing upward trends, images of bustling factories or thriving cities	Economic recovery is framed as a victorious battle; the visuals show prosperity and growth	Creates a sense of optimism and success, shaping public support for political leaders associated with recovery efforts
5. "Climate Crisis"	"Climate Change Crisis"	Images of natural disasters, melting glaciers, smoke from wildfires	The language emphasizes urgency, while visuals highlight the catastrophic nature of climate change	Increases public awareness of the environmental crisis, promoting action but also fear and helplessness about the future

TABLE 3. EXAMPLES OF MULTIMODAL REPRESENTATIONS IN MEDIA AND COMMUNICATION

Topic	Suggested Visual Products	Purpose of Visuals
1. "Illegal Alien"	- News article with the term "Illegal Alien" and a border fence background.	- Amplifies the criminalizing and dehumanizing nature of the term "illegal alien," associating immigrants with fear and danger.
	- Political cartoon showing immigrants as invaders, using alien imagery or military symbols.	- Reinforces the dehumanizing view of immigrants as invaders or "other."
	- Image of border fences with heavily patrolled areas, or migrants trying to cross.	- Frames immigration as a threat to national security.
2. "The Strongman"	- Photo of a political leader in front of national symbols, exuding confidence and authority.	- Visualizes the leader's power and dominance through nationalistic or militaristic imagery.
	- Leader speaking at a podium with military officers in the background, reinforcing strength.	- Reinforces the connection between power and authority with a militaristic backdrop.
	- Newspaper covers or political posters featuring the leader with iconic landmarks and intense lighting.	- Depicts the leader as the symbol of national strength and control.
3. "War on Drugs"	- Images of police raids, armed officers, and drug confiscation.	- Frames the drug issue as a war, promoting aggressive tactics for control.
	- Infographics showing rising arrests or spending on anti-drug programs.	- Symbolizes the ongoing battle against drugs through statistics and militarized approaches.
	- Photos of drug busts with armed officers and large amounts of confiscated substances.	- Visualizes the war on drugs with clear metaphors of conflict and law enforcement.
4. "Economic Recovery"	- Stock market charts showing positive growth or recovery.	- Reinforces the idea of successful economic recovery with visual data trends.
	- Photos of construction sites, new factories, or reopened businesses symbolizing job creation.	- Depicts revitalization and economic recovery through the growth of infrastructure and businesses.
	- Images of busy commercial areas, shoppers, and thriving neighborhoods symbolizing consumer confidence.	- Illustrates the return of prosperity and growth through active, bustling consumer spaces.
5. "Climate Crisis"	- Images of natural disasters such as wildfires, floods, or hurricanes, with people evacuating.	- Creates a sense of urgency and threat by visually linking

Topic	Suggested Visual Products	Purpose of Visuals
		climate change to destructive, life-threatening events.
	- Pictures of melting polar ice caps or coastal cities threatened by rising sea levels.	- Visualizes the tangible effects of climate change, emphasizing the crisis.
	- Infographics displaying temperature rise and impacts of climate change like extreme weather events.	- Condenses complex climate data into digestible visual formats, urging immediate action.
Additional Visual Products	**News Graphics**: Infographics combining language and visuals to convey data and trends.	- Simplifies complex issues and presents them in a visually appealing, easily understandable way.
	Political Cartoons: Caricatures and symbolic images combining humor and political commentary.	- Uses metaphor and humor to critique political figures and social issues, reinforcing ideologies.
	Public Service Announcements (PSAs): Government or NGO campaigns using language, imagery, and audio to communicate urgent messages.	- Combines various modes of communication to drive action on social issues like health, climate, and security.

Educational Approaches in English Philology

The recognition of multimodal communication holds profound implications for both English language teaching and various branches of English philology, such as historical linguistics, sociolinguistics, and translation studies. In the context of English philology, it is essential to move beyond a narrow focus on grammatical forms and engage with the broader semiotic resources that contribute to meaning-making, including visual, gestural, and prosodic cues. Language learners can benefit from an approach that incorporates these elements, as it offers a more holistic understanding of how language functions in real-world contexts.

For instance, in the teaching of metaphor, especially in English as a Foreign Language (EFL) or applied linguistics, the integration of visual prompts or spatial diagrams can significantly enhance learners' comprehension and retention of metaphorical concepts. Research by Boers and Lindstromberg (2008) has shown that presenting metaphors visually helps learners understand abstract concepts by relating them to familiar spatial or visual domains, thus strengthening cognitive connections. This approach is particularly relevant in the teaching of

conceptual metaphors in English, as it links linguistic structures with embodied experiences, creating richer learning opportunities.

In translation studies, a branch of applied philology, understanding the interplay between visual context and linguistic choices becomes crucial for achieving functional equivalence across languages and cultures. Multimodal translation, which accounts for both verbal and non-verbal elements, helps maintain the integrity of meaning in the translation process. For example, when translating advertisements or web content, attention to layout, typography, and visual images is necessary to preserve the intended impact of the original text. Thus, in translation practice, multimodal awareness ensures that meaning is communicated effectively, taking into account not only linguistic fidelity but also the non-verbal semiotic resources that influence interpretation.

Multimodal Corpus Approaches in Applied Philology

In contemporary applied philology, there has been a significant methodological expansion that reflects the broader multimodal turn in linguistic and semiotic research. Traditionally grounded in the close analysis of literary and historical texts, philology has increasingly embraced digital tools and empirical methods for the investigation of language in context. One of the most notable developments is the use of corpus tools not only for analysing linguistic patterns in written texts but also for engaging with multimodal data. Researchers now adapt and enrich corpora to include multimodal annotations, integrating layers of information that go beyond lexis and syntax to encompass prosody, gesture, spatial layout, color, font variation, and even interactional timing (Bateman et al., 2017; Baldry & Thibault, 2006).

This evolution in corpus design enables more nuanced investigations into how complex mental models and culturally situated ideologies are constructed and disseminated through multiple semiotic resources. For instance, multimodal corpora may annotate a politician's speech not only for lexical choices and syntactic structure, but also for gesture types (e.g., deictic or iconic), pitch modulation, gaze direction, and accompanying visual elements such as flags or background screens. Such detailed annotation allows for a richer interpretation of how political authority, emotional stance, and ideological frames are constructed and received (Cap, 2013; O'Halloran et al., 2014).

From a diachronic perspective, this approach also allows philologists to trace the historical development of multimodal meaning-making strategies in English across different epochs and media. For example, comparative studies between illustrated manuscripts, early printed pamphlets, and contemporary digital texts reveal how cultural narratives—such as national identity, morality, or gender roles—have

been visually and linguistically encoded in ways that are consistent with their respective communicative technologies and social contexts. The persistence of multimodal patterns, such as metaphorical motifs or recurring visual-linguistic alignments, across genres and time periods offers a fertile ground for exploring the interplay between language, cognition, and culture (Forceville, 2009; Ventola et al., 2004).

Moreover, these insights have significant pedagogical implications. In the classroom, integrating multimodal corpus approaches fosters a more inclusive and realistic understanding of English as it is used in real-world contexts. Students are no longer limited to studying literary texts or decontextualized grammar examples; instead, they can engage with authentic, multimodal materials such as news broadcasts, advertisements, or social media content. This shift aligns well with the educational aim of fostering critical digital literacy, allowing learners to deconstruct how meaning is shaped not just by what is said, but also by how it is presented visually and aurally (Jewitt, 2009; Serafini, 2014).

The integration of multimodal perspectives into English philology thus enriches both research and teaching. It foregrounds the importance of studying meaning as a distributed phenomenon—constructed through the dynamic interaction of verbal, visual, and embodied modes. By doing so, applied philology becomes better equipped to analyze the complexities of contemporary communication while also offering valuable tools for historical, literary, and cultural interpretation. This expanded methodological toolkit does not replace traditional philological practices but rather extends them, enabling scholars to address new questions and challenges in an increasingly multimodal and digitally mediated world.

Multimodality is not merely an additional layer to language; it is integral to the very structure of communication and essential for constructing and maintaining mental representations of reality. In the context of English, the combined use of semiotic modes—spoken and written language, gesture, images, sound, and spatial elements—forms a complex web through which we negotiate, share, and adapt our conceptual understandings of the world. These multimodal interactions are foundational to how we convey meaning, think about abstract concepts, and engage with others within specific cultural and social contexts.

Through the lens of multimodal analysis, English communication emerges as more than just a system of linguistic signs; it becomes a dynamic cognitive ecology where meaning is not isolated within the boundaries of language itself but is co-constructed through a range of interconnected modes. The significance of this approach goes beyond its ability to analyze how people communicate in real-world settings; it extends to how human cognition processes and organizes knowledge. As users of English, we rely on various modes to reinforce and complement one another, creating a cohesive framework that supports how we understand and express thoughts, ideas, and emotions.

This multimodal perspective is especially powerful in academic and practical fields such as applied linguistics, philology, translation studies, and discourse analysis, where a richer understanding of meaning-making can inform teaching, research, and professional practice. In English philology, for instance, recognizing the interdependence of verbal and non-verbal resources challenges traditional logocentric views that prioritize language alone and encourages a more comprehensive examination of how texts, genres, and communication systems operate. By integrating multimodal principles, scholars and educators can develop new methodologies that account for the multiplicity of semiotic resources at play in both historical and contemporary English texts.

Moreover, in applied philology and pedagogy, multimodal frameworks hold significant implications for teaching practices, particularly in how English is taught to non-native speakers. Instruction that emphasizes not just the grammatical structures of the language but also incorporates visual, auditory, and gestural elements fosters a more holistic understanding of communication. Whether through the teaching of metaphors, the analysis of visual culture in texts, or the application of multimodal translation strategies, this approach better prepares learners and professionals to navigate the complexity of meaning-making in real-world scenarios.

Ultimately, multimodality broadens the scope of English studies, opening new avenues for research that engage with the full spectrum of human communication. As both a theoretical paradigm and a practical framework, multimodality enriches our understanding of English not just as a system of signs but as a dynamic, cognitive environment that mirrors the full diversity of human experience. It underscores the interconnectedness of language, thought, and culture, offering fresh insights into the ways in which communication shapes and is shaped by the world around us. By incorporating this perspective into research and practice, scholars and practitioners can contribute to a deeper, more nuanced understanding of language and communication in the contemporary world.

SIGNIFICANCE FOR PHILOLOGY AND LINGUISTICS

The implications of multimodality for philological studies are profound. Historically, philology was concerned with the critical interpretation of texts, often focusing on issues such as authorship, manuscript tradition, and semantic shifts. While these remain central concerns, the incorporation of multimodal frameworks allows for a richer and more context-sensitive understanding of textual artifacts. For example, the layout, illustrations, and paratextual elements of medieval manuscripts

are now increasingly examined not merely as decorative additions, but as integral to the transmission and reception of meaning. This shift mirrors broader moves in literary studies and discourse analysis that foreground the material and multimodal nature of textuality.

In applied linguistics, multimodality has been embraced as a means to analyze communicative practices in a range of real-world contexts, from classroom interaction and advertising to digital media and political discourse. The theory enables researchers to understand how language functions in tandem with gesture, intonation, and spatial orientation in face-to-face communication, or how typography, color, and layout influence comprehension in digital texts. These insights are particularly relevant in English philology, where the analysis of evolving communicative forms—such as online discourse, visual narratives, and media-rich texts—requires tools that can account for more than just linguistic content.

THE IMPLICATIONS OF MULTIMODALITY FOR RESEARCH AND PEDAGOGY

The growing recognition of multimodality as a fundamental dimension of communication has profound implications for both academic research and educational practice. Multimodal theory challenges the long-standing dominance of verbal and written language as the primary conveyors of meaning by recognizing the integral role of other semiotic resources—such as images, gesture, sound, color, spatial layout, and movement—in shaping understanding and experience (Kress & van Leeuwen, 2001; Jewitt, 2009). As a result, multimodal approaches offer enriched frameworks for investigating how meaning is constructed, negotiated, and interpreted across a variety of communicative contexts.

In the field of pedagogy, multimodal approaches are particularly transformative. Language educators, especially those involved in second or foreign language instruction, increasingly draw on multimodal strategies to enhance comprehension, engagement, and retention. Visual aids such as diagrams, infographics, and illustrated vocabulary support learners by providing context and reducing cognitive load (Mayer, 2005). Gestures, facial expressions, and prosody help make abstract or unfamiliar concepts more concrete, especially for learners with limited linguistic proficiency. For example, in a classroom discussion about environmental issues, a teacher may pair the word "deforestation" with images of barren landscapes and gestures mimicking tree-felling, reinforcing understanding through multiple sensory channels. This aligns with embodied and multimodal learning theories that stress the importance of integrating sensory, cognitive, and emotional modes of engagement (Flewitt et al., 2014; Bezemer & Kress, 2016).

Moreover, multimodal pedagogy encourages active student participation and production. Assignments such as creating digital narratives, video essays, or multimodal presentations allow learners to express themselves through combinations of text, image, sound, and interactivity. This not only fosters creativity but also develops students' critical awareness of how different modes function rhetorically—how, for example, font size, color, and visual framing can influence an audience's interpretation just as much as word choice can (Serafini, 2014). Such competencies are increasingly vital in a media-saturated world, where the ability to analyze and produce multimodal content is central to digital literacy.

In literary and cultural studies, the implications of multimodality are equally significant. Scholars are re-evaluating genres such as graphic novels, films, and digital narratives as sophisticated, layered forms of storytelling that demand multimodal analysis. Rather than viewing images as mere illustrations of text, multimodal theory allows for an exploration of how text and image co-construct meaning in dynamic and sometimes conflicting ways. For instance, in graphic novels like *Persepolis* or *Maus*, visual symbolism, panel layout, typography, and spatial pacing contribute crucial narrative and emotional information that cannot be captured through linguistic analysis alone (Groensteen, 2007; McCloud, 1993). Similarly, in film studies, multimodal frameworks enable scholars to analyze how narrative progression is shaped not just by dialogue and plot but by cinematography, sound design, color grading, and visual metaphor (Baldry & Thibault, 2006; Bateman & Schmidt, 2012).

Multimodality also opens new avenues for historical research, particularly in the study of textual artifacts. In manuscript studies, for instance, features such as illumination, marginalia, typography, and spatial organization have traditionally been treated as decorative or peripheral. However, from a multimodal perspective, these features are understood as integral components of meaning-making. A medieval manuscript is not simply a vessel for transcribed words—it is a richly multimodal object, with every visual and material choice (from parchment texture to ink color) contributing to how the text was experienced and interpreted by its original readers (Jäger, 2015). Similarly, early printed texts used typography, woodcut illustrations, and layout to guide reader attention, create emphasis, or signal genre, much as contemporary digital interfaces do. These insights call for a reassessment of long-standing hermeneutic traditions in the humanities, urging scholars to consider how meaning is distributed across modes and materialities rather than being confined to linguistic content.

Crucially, multimodality also intersects with questions of ideology and representation. In both contemporary and historical texts, visual and spatial arrangements often carry ideological weight. In newspapers, for example, headline placement, image selection, and font size all contribute to framing effects that influence reader perception of events (Machin & Mayr, 2012). In religious or political texts, symbolic imagery

and ornamental design may function to legitimize authority or convey spiritual truths. As a result, multimodal analysis provides powerful tools for uncovering the subtle mechanisms through which power, identity, and belief are communicated and contested across cultures and epochs.

In sum, multimodality reshapes the ways in which we approach language, literature, education, and history. It provides a framework for understanding meaning as a product of interrelated semiotic resources, each contributing uniquely to communication. For researchers and educators alike, the adoption of multimodal perspectives enhances interpretive depth, pedagogical effectiveness, and critical literacy. As such, multimodality is not merely an auxiliary concern but a central paradigm for 21st-century scholarship.

MULTIMODALITY AND ITS IMPACT ON ENGLISH PHILOLOGY AND DISCOURSE STUDIES: A NEW PARADIGM FOR EXPRESSION AND COMPREHENSION

The interdisciplinary nature of multimodality brings together various fields of study, offering a revolutionary shift in how meaning is produced, interpreted, and understood. In English philology, traditionally focused on linguistic structures, meaning-making processes, and textual analysis, the integration of multimodal frameworks has introduced novel ways to explore both historical and contemporary texts. This approach challenges long-standing paradigms by recognizing that language and meaning are not just confined to the spoken or written word but are enriched by a combination of semiotic resources, including gesture, image, sound, and layout. Thus, multimodality provides both a theoretical and methodological innovation for philological research.

In traditional philological studies, the focus has typically been on written texts, emphasizing syntax, semantics, and literary devices within a linguistic framework. However, in an increasingly visual and multimedia-driven world, this limited approach fails to account for the ways in which communication is inherently multimodal. English philology, as it is currently understood, has primarily been concerned with language as a system of signs that can be deconstructed, analyzed, and interpreted through the lens of textual and verbal analysis. However, this paradigm does not fully capture the depth of human communication, which involves far more than just words. Multimodal analysis opens up new avenues for exploring how different semiotic modes interact to create meaning, and this shift invites philology to rethink its boundaries.

One of the key contributions of multimodal theory to philology is its emphasis on the integration of cognitive linguistics. Cognitive linguistics posits that meaning is shaped by the mind's conceptual structures and that these structures are influenced by sensory experiences, including sight, sound, and touch. In multimodal communication, various sensory channels come into play, and the meaning is not solely generated through linguistic input. This aligns with the cognitive turn in linguistics, where scholars argue that understanding language requires a focus on how humans mentally conceptualize their surroundings and how different semiotic modes—such as images, gestures, and even space—contribute to the construction of meaning.

This interdisciplinary perspective on communication opens the door to more comprehensive approaches in philological research. The study of manuscripts, for example, no longer focuses solely on the written word or text's linguistic properties but also includes an examination of physical attributes like layout, handwriting style, and even illustrations or marginalia. These elements are not mere accessories to the text but are integral to how the meaning of a document is conveyed and understood. In this way, multimodal philology extends beyond the limits of traditional philological analysis, encouraging scholars to consider the full spectrum of semiotic resources that shape the meaning-making process.

The practical applications of multimodality extend to a wide range of disciplines within English philology. For instance, in the field of translation studies, it is no longer enough to simply consider the linguistic equivalence between two languages. A multimodal approach to translation requires understanding how visual cues, sounds, and even spatial arrangements interact with verbal language to produce meaning in both the source and target cultures. Similarly, in literary studies, multimodal analysis can reveal how authors construct meaning not only through language but also through images, symbols, and spatial arrangements, whether in printed texts, film adaptations, or digital media.

The rise of digital communication has particularly underscored the relevance of multimodal approaches. In online environments, communication is often conveyed through a combination of written words, images, videos, hyperlinks, and even interactive elements. As a result, English philology must adapt to this new landscape, where the traditional boundaries of written and spoken language are often blurred. Digital texts—whether social media posts, websites, or online advertisements—rely heavily on multimodal elements to engage the reader, convey meaning, and influence behavior. Thus, applying a multimodal framework to the study of digital texts is crucial to understanding how contemporary English is used to construct identity, shape ideologies, and communicate across cultures.

From a pedagogical perspective, the integration of multimodal theory into English philology holds great promise for the future of language teaching. Traditional language instruction has focused on grammar, vocabulary, and reading comprehension, with less emphasis on the visual, spatial, and cultural contexts in which language is used. A multimodal approach to language teaching expands this view by emphasizing how students interact with and interpret different semiotic modes. For example, teaching metaphor can be enhanced by providing visual prompts or spatial diagrams that help students understand how metaphors operate across different modes of communication. Similarly, students of translation can learn not only to translate words but to navigate the complex interplay of verbal and visual elements across languages and cultures.

The growing significance of multimodal communication in our globalized world underscores the importance of this interdisciplinary framework for English philology. As new technologies continue to emerge, scholars and educators must remain attuned to the evolving modes of communication, ensuring that the study of English is responsive to the complex, multimodal reality in which we live. By embracing the insights of cognitive linguistics, multimodal analysis, and applied philology, English philology can be redefined as a discipline that fully accounts for the richness and complexity of human communication.

In conclusion, multimodality represents a transformative paradigm for English philology, offering a more holistic understanding of how language and meaning are constructed. By integrating cognitive linguistics and multimodal theory into philological research, scholars can gain deeper insights into both historical and contemporary texts. Furthermore, this approach allows for a more nuanced appreciation of how language functions in a world where communication is increasingly shaped by diverse semiotic resources. As the study of multimodal communication continues to grow, it will undoubtedly lead to new perspectives and methodologies that will shape the future of English philology, enabling it to respond to the challenges and opportunities of the modern world.

POLITICAL DISCOURSE AND COGNITION: UNVEILING THE POWER OF LANGUAGE IN SHAPING POLITICAL THOUGHT

Political discourse represents a fertile ground for the study of how language influences thought, opinion, and action, particularly through the use of metaphors. Metaphors in political language do much more than simply describe issues—they actively shape perceptions, evoke strong

emotions, and often serve as tools for framing political ideologies in ways that can manipulate public opinion. By studying political discourse through the lens of cognitive linguistics, we gain a deeper understanding of how metaphors are not just rhetorical flourishes, but fundamental mechanisms that construct and reinforce political ideologies.

One of the most significant contributions of cognitive linguistics to political analysis is its focus on the conceptual nature of metaphors. Cognitive linguists argue that metaphors are not merely linguistic expressions but are reflections of the conceptual frameworks that shape the way we think about the world. In political discourse, metaphors such as "nation-as-a-family" or "immigration-as-a-flood" are prime examples of how political actors frame complex social and political issues in ways that evoke specific emotional responses and align with particular ideological positions. These metaphors are not innocent linguistic tools—they are laden with powerful cognitive effects that can shape how the public understands and responds to political issues.

The cognitive dimension of political metaphors becomes particularly clear when we examine how these metaphors activate mental models that reinforce specific worldviews. The theory of conceptual metaphors, proposed by Lakoff and others, suggests that metaphors shape our cognitive structures by providing a mental framework through which we interpret and make sense of the world. In the context of political discourse, these metaphors often serve to create distinct mental models that are deeply ingrained in political ideologies. For instance, conservatives in the U.S. tend to favor metaphors that reflect a “strict father” model, where the state is seen as an authority figure that enforces order and discipline. This metaphorical framework evokes a worldview centered on hierarchy, personal responsibility, and a belief in self-reliance.

In contrast, liberals tend to prefer metaphors associated with a “nurturant parent” model, where the state is viewed as a protector and caregiver, responsible for nurturing and supporting citizens, especially those in need. This metaphorical framing evokes a different set of values, emphasizing community, care, and social responsibility. These divergent metaphorical systems are not merely abstract concepts—they resonate with deeply held beliefs and provide the cognitive tools through which individuals interpret political debates, social policies, and national identity. The metaphors of “strict father” and “nurturant parent” illustrate how metaphors serve as cognitive mechanisms that can activate mental models aligning with specific ideological perspectives.

The power of metaphorical framing in political discourse lies in its ability to reinforce these mental models, solidifying ideological divides. Political rhetoric often employs metaphors strategically, not just to describe or clarify issues, but to subtly guide how the audience should

think about them. By framing an issue in one metaphorical context, political leaders can steer public perception in a way that favors their own ideological stance. For instance, the use of the "war on drugs" metaphor activates a mental model where drug use is seen as an enemy to be fought, while the “public health crisis” metaphor might frame the issue as a societal problem requiring compassionate solutions.

Moreover, metaphors are not only used to evoke particular values and emotions, but they also serve to construct political identities and alignments. The metaphors employed in political discourse can serve as boundary markers, distinguishing one political group from another. Through metaphoric framing, political leaders can create a sense of "us" versus "them," further deepening ideological divides. For example, the metaphor of the “American Dream” has been central to conservative ideologies, evoking a vision of individual success through hard work and personal responsibility. On the other hand, liberal political discourse often invokes metaphors of “community” and “shared responsibility,” calling for collective action to address societal inequities.

The role of metaphors in political discourse also highlights the connection between language and power. Metaphors have the ability to shape public opinion in subtle yet profound ways. Political elites and media organizations are keenly aware of the power of metaphor and often use it as a tool to influence public opinion and shape the political landscape. By controlling the metaphors used in public discourse, political actors can shape how citizens think about key issues, often without their awareness. This makes the study of political metaphor not only an exploration of linguistic creativity but also an important avenue for understanding the mechanisms of political power and manipulation.

Additionally, the cognitive-linguistic approach to political metaphor reveals the dynamic interplay between language and thought. While metaphors are shaped by cultural and political contexts, they also, in turn, shape the way we perceive the world and influence our decision-making. Through the repeated use of certain metaphors in political discourse, these conceptual frameworks become entrenched in the public consciousness, shaping political attitudes and behaviors for generations. As a result, understanding the metaphors that dominate political discourse is key to understanding how public opinion is formed and how ideological divides are perpetuated.

The use of metaphors in political discourse is not a mere rhetorical technique; it is a powerful cognitive tool that shapes political thought, constructs worldviews, and reinforces ideological divides. Cognitive linguistics provides valuable insights into the ways in which metaphors activate mental models that align with specific political ideologies. By studying political metaphors, we uncover the deep cognitive structures that underlie political discourse and gain a better understanding of how

language plays a central role in shaping public opinion and political behavior. In an age of increasing polarization, the analysis of metaphors offers an essential lens through which we can examine the forces that shape political life.

The intersection between Cognitive Linguistics and Multimodality offers a particularly rich area of inquiry, as both frameworks emphasize the embodied, usage-based, and meaning-driven nature of communication. While Cognitive Linguistics traditionally focuses on verbal language and its cognitive underpinnings, Multimodality expands this perspective by investigating how meaning is constructed across multiple semiotic modes, including gesture, visual imagery, spatial layout, sound, typography, and other non-linguistic resources (Kress & van Leeuwen, 2001).

Multimodality shares Cognitive Linguistics' fundamental commitment to the idea that communication is not solely linguistic, but deeply contextual, embodied, and conceptual. For instance, just as CL posits that metaphors are grounded in physical experience (e.g., the *UP* is *GOOD* metaphor), multimodal theory examines how such metaphors are also realized visually—through upward movement in charts, or elevated placement of images to indicate importance. In this way, the conceptual metaphors studied in Cognitive Linguistics are not only expressed in spoken or written language, but also visually and spatially, often reinforcing or expanding meaning across modes (Forceville & Urios-Aparisi, 2009).

In practical terms, the integration of CL and multimodal approaches is particularly evident in fields like translation studies, discourse analysis, and educational linguistics, where researchers analyze how meaning is conveyed and transformed across linguistic and visual systems. For example, in audiovisual translation (AVT), understanding how subtitles (verbal), body language (visual), and intonation (auditory) contribute to meaning aligns with both CL's focus on conceptualization and multimodality's attention to semiotic orchestration. Similarly, multimodal metaphors—where one element of the metaphor is presented linguistically and another visually—require a cognitive approach to decoding meaning across formats.

Moreover, the study of terminology, particularly in technical and scientific domains, increasingly acknowledges the multimodal nature of knowledge communication. Diagrams, charts, color-coded taxonomies, and iconographic systems often complement or even substitute verbal definitions in specialist discourse. Here, CL's concern with how conceptual structures are organized in the mind pairs well with multimodal analysis of how those structures are externalized and communicated through visual and spatial means. For instance, a term like *feedback loop* in environmental science might be understood more fully

through both a verbal explanation and a visual representation, each mode contributing distinct but complementary aspects of meaning.

Thus, bridging Cognitive Linguistics with Multimodality allows for a more comprehensive understanding of how meaning is created and interpreted in real-world contexts. Both theories challenge reductive, language-only models of communication and instead promote an integrative view where language, thought, and perception interact across sensory and semiotic channels. As communication becomes increasingly mediated through digital platforms, visual media, and interdisciplinary formats, this intersection will continue to be of growing relevance for linguistic research, terminology development, and educational practice.

References

Adami, E. (2015). What's in a click? A social semiotic framework for the multimodal analysis of website interactivity. *Visual Communication*, *14*(2), 133–153.

Baldry, A., & Thibault, P. J. (2006). *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook*. Equinox.

Barros, R., & Garcia, L. (2019). *Social Media and Multimodal Communication*. Springer.

Bateman, J., & Schmidt, K. H. (2012). *Multimodal Film Analysis: How Films Mean*. Routledge.

Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodal Analysis in the Age of Digital Communication*. Routledge.

Bell, A. (1991). *The Language of News Media*. Blackwell.

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166–195.

Bezemer, J., & Kress, G. (2016). *Multimodality, Learning and Communication: A Social Semiotic Frame*. Routledge.

Cap, P. (2013). *Proximization: The Pragmatics of Political Discourse*. John Benjamins.

Charteris-Black, J. (2011). *Politicians and Rhetoric: The Persuasive Power of Metaphor* (2nd ed.). Palgrave Macmillan.

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Danesi, M. (2016). *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. Bloomsbury.

Davies, M. (2008). *The Corpus of Contemporary American English: 450 Million Words, 1990–Present*. Retrieved from https://www.english-corpora.org/coca/

Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. Lawrence Erlbaum Associates.

Flewitt, R., Hampel, R., Hauck, M., & Lancaster, L. (2014). What are multimodal data and transcription? In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis* (2nd ed., pp. 44–59). Routledge.

Forceville, C. (2009). Non-verbal and multimodal metaphor in a

cognitivist framework: Agendas for research. In C. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor* (pp. 19–42). Mouton de Gruyter.
Forceville, C., & Urios-Aparisi, E. (Eds.). (2009). *Multimodal Metaphor*. Mouton de Gruyter.
Greenbaum, S., & Nelson, G. (1996). *The International Corpus of English: ICE-GB*. Oxford University Press.
Groensteen, T. (2007). *The System of Comics*. University Press of Mississippi.
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge.
Jäger, L. (2015). *Introduction to the Theory of Multimodality*. De Gruyter.
Jewitt, C. (2009). *The Routledge Handbook of Multimodal Analysis*. Routledge.
Jewitt, C., & Kress, G. (2003). *Multimodal Literacies: A Handbook for Researching and Teaching Communication*. Peter Lang.
Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing Multimodality*. Routledge.
Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.
Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge.
Kress, G., & van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. Arnold.
Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd ed.). Routledge.
Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press.
Lakoff, G. (2004). *Don't Think of an Elephant: Know Your Values and Frame the Debate*. Chelsea Green Publishing.
Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.
Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. Stanford University Press.
Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
Leech, G., Conrad, S., & Scott, P. (2001). *The British National Corpus: A Guide to the Design and Use of the Corpus*. Oxford University Press.
Mahlberg, M. (2018). *Corpus Linguistics and the Study of Transitivity in Literature: Insights from CLiC*. University of Birmingham Press.
Machin, D., & Mayr, A. (2012). *How to Do Critical Discourse Analysis: A Multimodal Introduction*. SAGE.
Mayer, R. E. (2005). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press.
McCloud, S. (1993). *Understanding Comics: The Invisible Art*. HarperCollins.
McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182. https://doi.org/10.1126/science.1199644
Page, R. (2018). *Narratives Online: Shared Stories in Social Media*. Cambridge University Press.
Serafini, F. (2014). *Reading the Visual: An Introduction to Teaching Multimodal Literacy*. Teachers College Press.
Talmy, L. (2000). *Toward a Cognitive Semantics, Volume I: Concept Structuring Systems*. MIT Press.
Ventola, E., Charles, C., & Kaltenbacher, M. (Eds.). (2004). *Perspectives on Multimodality*. John Benjamins.
Wertheimer, M. (1923). Laws of organization in perceptual forms. *Psychologische Forschung, 4*(1), 301–350.
Zappavigna, M. (2012). *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web*. Continuum.

PART III

GRAMMAR AS MULTIMODAL MEANING-MAKING

Traditional linguistic theories have long positioned grammar as an autonomous, rule-governed system concerned primarily with the structure of language—particularly syntax and morphology. Within this paradigm, meaning is often considered secondary to form, and the role of language is delimited to linear, textual expression. However, recent developments in multimodal theory challenge this bounded view by proposing that grammar is not limited to verbal forms but is inherently intertwined with other semiotic resources such as gesture, prosody, visual layout, and spatial orientation. These resources do not merely accompany language; they actively participate in the grammatical structuring of meaning.

In spoken interaction, for instance, intonation and rhythm—traditionally analyzed under prosody—play an essential role in demarcating clause boundaries, identifying thematic progression, and distinguishing sentence types (Halliday & Matthiessen, 2014). Similarly, deictic gestures such as pointing can function analogously to grammatical elements like demonstratives or spatial adverbs, co-constructing reference and anchoring utterances in the immediate physical context (McNeill, 2005). In digital and visual communication, typography, spacing, and layout frequently fulfill grammatical functions by signaling syntactic divisions or emphasizing textual hierarchies, thereby performing a role akin to punctuation in written language (Ledin & Machin, 2016).

Building on these insights, scholars such as Christian M.I.M. Matthiessen and Theo van Leeuwen have proposed models of multimodal or visual grammar that extend traditional grammatical frameworks into non-verbal modalities. Van Leeuwen (2006), for example, adapts linguistic concepts such as modality, transitivity, and salience to the analysis of images, demonstrating that visual texts exhibit systematic, rule-governed structures comparable to those found in verbal grammar. These approaches rest on the premise that different modes possess their own "grammars"—semiotic systems with internal logics that govern how meaning is constructed and interpreted across modalities.

From the perspective of cognitive and embodied linguistics, such developments are not merely extensions of grammar but a rethinking of its foundations. Cognitive Grammar, as articulated by Langacker (2008), posits that grammatical structures emerge from usage and are grounded in embodied experience and perceptual patterns. Language, in this view, is not an autonomous symbolic code but an expression of general cognitive capacities shaped by interaction with the physical and social world. Multimodal elements—gesture, image, sound—are likewise rooted in perception and action, reflecting the same embodied schemata

that underlie grammatical categories. Thus, grammar itself can be seen as a multimodal phenomenon, distributed across the body, voice, and environment, and dynamically enacted in context.

This reconceptualization invites us to move beyond the notion of grammar as a static set of rules and toward an understanding of grammar as an emergent, multimodal process of meaning-making. It aligns linguistic theory with contemporary understandings of communication as inherently multimodal, situated, and embodied, offering a richer and more integrated model of how humans produce and interpret meaning.

Toward a Multimodal Reconceptualization of Theoretical Grammar

In light of recent advances in multimodal and cognitive linguistic theory, the traditional architecture of theoretical grammar is increasingly in need of critical revision. Conventional grammar courses, grounded primarily in structuralist or generativist paradigms, continue to conceptualize language as an autonomous, rule-governed system largely insulated from other forms of meaning-making. This model, although analytically rigorous, abstracts language from the dynamic, situated contexts in which communication actually unfolds. It overlooks the fact that human meaning-making is inherently multimodal, embodied, and socially distributed (Kress, 2010; Jewitt, Bezemer, & O'Halloran, 2016). To address this gap, an expanded grammar curriculum informed by Multimodal Metaphor Theory (Forceville & Urios-Aparisi, 2009) could offer a more comprehensive and realistic account of how grammar operates within broader semiotic ecologies.

Reconceptualizing grammar through a multimodal lens involves rethinking it not as a closed, autonomous system but as one component within a network of interrelated semiotic modes—including gesture, gaze, image, prosody, spatial arrangement, and kinesthetic interaction. These modalities do not merely embellish or support language; they are constitutive of grammar itself in many communicative contexts. For example, gesture can perform syntactic functions, serving as deixis or reinforcing transitivity patterns (McNeill, 2005; Kendon, 2004). Intonation and rhythm are not external to grammatical form but are instrumental in establishing clausal boundaries and highlighting information structure (Wichmann, 2000). In digital environments, typographic and visual design choices—such as bolding, spacing, and alignment—operate grammatically to indicate emphasis, hierarchy, or segmentation, thereby shaping reader interpretation in ways comparable to traditional punctuation (Ledin & Machin, 2016).

Despite these insights, traditional grammar instruction remains largely monomodal, with little attention paid to the multimodal nature of real-world language use. This omission is not merely a pedagogical oversight but a theoretical limitation. By excluding non-verbal modalities, current models of grammar risk presenting an impoverished and idealized version of language, one that fails to account for the embodied and

interactive nature of communication. Incorporating concepts from Multimodal Metaphor Theory could help redress this imbalance. MMT demonstrates how metaphor operates not only in linguistic forms but across modes—through images, gestures, sound, and spatial positioning—revealing metaphor as a cross-modal, experiential phenomenon (Forceville, 2007; Cienki & Müller, 2008). Grammar, in turn, can be understood as both supporting and emerging from such metaphorical mappings, particularly when viewed through the lens of Cognitive Grammar, which sees grammatical structure as semantically motivated and experientially grounded (Langacker, 2008).

A revised grammar curriculum might therefore benefit from thematic modules that focus on multimodal syntax, embodied grammatical forms, and cognitive-discursive integration. Students could explore how grammatical meaning is distributed across facial expression, prosody, and gesture; how image schemas and conceptual metaphors shape grammatical categories such as aspect or tense; and how grammatical structure adapts in digital and visual media forms, including memes, interactive interfaces, and cinematic discourse (Bateman, Wildfeuer, & Hiippala, 2017). These additions would not displace traditional grammatical knowledge but would contextualize it within a more ecologically valid framework of multimodal communication.

However, while several scholars have called for multimodal extensions to grammar (e.g., Matthiessen, 2007; van Leeuwen, 2005), few have attempted to systematize such insights into a comprehensive pedagogical model. Theoretical development remains fragmented, and a coherent synthesis of multimodal grammar is still lacking. Moreover, most models have focused on specific modes (e.g., visual grammar) without fully integrating them into a general grammar theory that includes metaphor and cognition. This represents both a gap and an opportunity.

To capture this emerging perspective, it may be necessary to articulate a new theoretical stage in the development of Multimodal Metaphor Theory. Possible designations include *Multimodal Grammar Theory*, which emphasizes grammar as inherently multimodal in structure and function; *Embodied Multimodal Metaphor Theory*, which underscores the role of sensorimotor experience in cross-modal mapping; or *Integrated Multimodal Meaning Theory*, which aims to synthesize grammatical, metaphorical, and semiotic dimensions into a unified framework of meaning-making. Each of these formulations seeks to move beyond merely additive models—where gesture or image is treated as supplementary—and toward a truly integrative view of communication where grammar is reconceived as a distributed, multimodal, and embodied process.

IMPLICATIONS OF MULTIMODAL FRAMEWORKS IN ENGLISH PHILOLOGY AND LANGUAGE PEDAGOGY: EXPANDING THE SCOPE OF TEXTUAL ANALYSIS AND INSTRUCTION

In English philology and language pedagogy, multimodal frameworks challenge traditional logocentric assumptions by highlighting the role of non-verbal modes of communication in the construction and interpretation of meaning. Traditionally, philology has focused primarily on written texts and linguistic forms, often excluding the various semiotic resources that shape how meaning is conveyed in real-world communication. However, by integrating multimodal analysis, philology can now be expanded to consider the embodied and visual dimensions of meaning-making, thus offering a more holistic approach to textual analysis. For instance, when examining literary texts, scholars can move beyond mere syntactic structures and explore how visual layout, gestures, and even the spatial positioning of text in relation to illustrations or other media elements contribute to the overall meaning. This shift in perspective is not only beneficial for understanding historical texts, such as illuminated manuscripts or early printed books, but it also has significant implications for contemporary textuality, which increasingly involves digital and multimodal communication (Bateman, Wildfeuer, & Hiippala, 2017).

In the context of language pedagogy, particularly in teaching English as a second or foreign language, multimodal frameworks provide powerful pedagogical tools. Language learning is often seen as a cognitive process primarily dependent on linguistic structures; however, by incorporating non-verbal elements such as gestures, visual aids, and digital media, learners' understanding of grammatical concepts can be significantly enhanced. For example, gesture-based instruction can make abstract grammatical structures more tangible by linking them to physical actions, which in turn anchors learners' comprehension in embodied experience (Kendon, 2004). The use of visual aids—such as infographics, diagrams, and video clips—further supports learners by providing additional cognitive cues that facilitate the internalization of linguistic rules. In this way, multimodal resources do not only reinforce grammatical structures but also offer alternative modes of expression that can accommodate diverse learning styles and needs.

Such pedagogical practices align with the principles of multimodal literacy, which emphasize the integration of various semiotic modes—such as language, gesture, images, and sound—in communication (Kress, 2010; Jewitt et al., 2016). In this paradigm, language is understood not as an isolated system but as part of a broader, multimodal communicative ecology. Teaching English through multimodal means encourages learners to engage with language in a more dynamic and contextually situated manner. Rather than learning grammatical rules in isolation, students are exposed to the ways in which grammar functions

in concert with other semiotic resources in real-world communication. This approach mirrors how language is used in daily life, where meaning is co-constructed through an interplay of spoken and written words, visual elements, and bodily gestures. As such, multimodal teaching practices contribute to a more holistic educational paradigm that better reflects the complexity of human communication.

Moreover, by broadening the scope of language instruction to include multimodal resources, teachers foster critical thinking and awareness about the role of visual and auditory modes in meaning-making. Students become not only more proficient in grammar but also more attuned to the ways in which different semiotic modes interact and reinforce each other in various communicative contexts. This holistic approach is particularly relevant in an increasingly globalized and digital world, where communication is rarely confined to one mode and often requires the simultaneous processing of multiple forms of input. The integration of multimodal frameworks into pedagogy thus equips learners with the skills necessary to navigate and interpret a wide range of communicative scenarios, from face-to-face interactions to digital discourse, where meaning is constructed through the collaboration of text, image, sound, and gesture.

Ultimately, the adoption of multimodal frameworks in both philology and pedagogy offers an opportunity to move beyond the limitations of logocentric models of language and communication. By considering the full spectrum of semiotic resources available for meaning-making, scholars and educators can gain a more nuanced understanding of how language operates in real-world contexts. As multimodal approaches continue to gain traction, they promise to reshape the way we study, teach, and understand English, fostering a deeper connection between linguistic theory and the diverse modes of communication that shape human interaction in the contemporary world.

Expanding the Scope of Textual Analysis and Instruction: A Multimodal Approach

The evolution of communication, both in theory and practice, necessitates a shift in how we approach the analysis and instruction of language. In traditional models of philology and pedagogy, language is often treated as an isolated system of signs, focused primarily on verbal or written communication. However, the rise of multimodal communication—where meaning is conveyed through a combination of linguistic, visual, auditory, and gestural elements—compels us to rethink this framework. By expanding the scope of textual analysis and instruction to encompass these diverse semiotic modes, scholars and educators can foster a more nuanced understanding of how meaning is constructed in real-world settings. This shift challenges the logocentric view of language and opens up new avenues for exploring how language interacts with other forms of expression in both historical and contemporary contexts.

In the realm of English philology, a multimodal perspective offers the potential to expand our understanding of texts beyond the limits of written or spoken words. Traditionally, philology has been concerned with the study of ancient manuscripts, literary works, and linguistic structures within specific cultural and historical contexts. However, this narrow focus often neglects the interplay between text and other semiotic resources, such as images, gesture, and spatial organization. For instance, illuminated manuscripts, which combine written text and visual elements, demand a more integrated approach that considers both the linguistic content and the visual layout. The same holds true for contemporary digital texts, which frequently incorporate images, videos, hyperlinks, and other multimodal components that influence the overall meaning. A multimodal framework in philology, therefore, invites scholars to consider how these various modes work together to create meaning, rather than treating them as separate, isolated elements. This broader approach allows for a more comprehensive understanding of how texts function in their cultural and historical contexts, as well as how meaning is negotiated through the interaction of multiple semiotic systems.

Similarly, in language pedagogy, particularly in teaching English as a second or foreign language, a multimodal approach can transform traditional teaching methodologies. Historically, language instruction has focused primarily on grammar, vocabulary, and pronunciation, with an emphasis on written and spoken language. While these aspects are undeniably important, they represent only one facet of communication. The integration of multimodal resources—such as gesture, visual aids, and digital media—offers a richer, more embodied approach to language learning. By incorporating gestures and body language into language instruction, for example, learners can connect abstract grammatical concepts to physical actions, making them more tangible and easier to understand. This is particularly beneficial for second language learners, who may struggle to grasp grammatical structures through verbal explanations alone. Gesture, as an embodied mode of communication, not only enhances learners' understanding of grammar but also fosters a deeper connection to the language itself, reinforcing the idea that meaning is not simply an intellectual exercise but a dynamic, embodied process (Kendon, 2004).

Visual aids and digital media further enhance language learning by providing learners with additional cognitive cues that reinforce grammatical concepts. For instance, infographics, diagrams, and videos can be used to illustrate the structure of complex grammatical rules, such as sentence syntax or tense usage. This multimodal approach aligns with the principles of multimodal literacy, which emphasizes the integration of various semiotic modes—text, image, sound, and gesture—in communication (Kress, 2010; Jewitt et al., 2016). By incorporating these modes into language instruction, educators can help students understand that meaning is not confined to the linguistic system alone but is co-constructed through a variety of semiotic resources. This broader

perspective encourages learners to engage with language in a more dynamic and contextually situated manner, preparing them to navigate the complexities of real-world communication, where meaning is often conveyed through a combination of modes.

In addition to enhancing grammatical understanding, the multimodal approach also fosters a deeper level of engagement with the material. Traditional language instruction often treats language as a set of rules and structures to be memorized, whereas multimodal approaches encourage students to actively engage with language in diverse and meaningful ways. For example, in the study of metaphors, learners can explore how metaphorical concepts are not only expressed through words but also represented in visual media, such as advertisements, films, and social media posts. By analysing how metaphors function across different modes, students can develop a more comprehensive understanding of how abstract concepts are constructed and communicated. This not only deepens their knowledge of language but also enhances their ability to analyze and produce multimodal texts in their own work.

Moreover, the integration of multimodal frameworks into language pedagogy supports the development of critical thinking skills. Students are encouraged to analyze how different semiotic modes—such as text, image, and sound—interact and reinforce each other in the communication process. This critical engagement with multimodal texts helps students understand that meaning is not fixed or inherent in any one mode but is constructed through the dynamic interplay of various resources. By fostering this awareness, educators can help students become more critical consumers of media and more effective communicators in a world where meaning is often conveyed through multiple channels. This is particularly important in the digital age, where communication is rarely confined to one mode and often involves the simultaneous processing of text, image, video, and sound. By equipping students with the tools to analyze and produce multimodal texts, educators prepare them to participate fully in the complex, multimodal communicative environments of the contemporary world.

The benefits of a multimodal approach to language instruction extend beyond the classroom and have significant implications for real-world communication. As communication increasingly occurs through digital platforms—such as social media, websites, and online forums—individuals must be able to navigate and interpret multimodal texts in order to engage effectively with others. In these contexts, meaning is not solely constructed through language but is co-constructed through the interaction of text, image, sound, and gesture. By teaching students to analyze and produce multimodal texts, educators provide them with the skills necessary to participate in these digital spaces and engage with the full range of communication modes available. This ability to understand and navigate multimodal texts is increasingly essential for success in the modern world, where effective communication often requires the ability

to synthesize multiple forms of information and respond in an appropriate and meaningful way.

To advance the integration of Multimodal Theory (MT) into textual analysis, a more systematic and empirically grounded approach is needed—one that explores not only the theoretical foundations but also the specific mechanisms by which various semiotic modes contribute to meaning-making. This area remains underdeveloped, particularly in terms of applying multimodal frameworks across diverse corpora and communicative contexts.

Multimodal textual analysis requires reframing the notion of "text" as a multimodal construct, where linguistic expression is intricately interwoven with gesture, image, sound, spatial arrangement, and embodied interaction. However, current treatments often remain at the illustrative or conceptual level, and more sustained, corpus-driven research is necessary to clarify how these modes interact systematically across genres, registers, and cultural practices. While studies in multimodal discourse analysis (e.g., Kress & van Leeuwen, 2006; Bateman, 2014) have made important advances, they have yet to fully address the empirical diversity of communicative practices—ranging from political discourse to language classrooms, from advertising to digital storytelling.

The identification and categorization of modalities and their functional contributions to grammar and rhetoric demand closer scrutiny through annotated multimodal corpora. Such corpora would allow researchers to track recurring intermodal patterns, such as how gestures reinforce deictic reference in spoken discourse, or how image-schema mappings in visual metaphors align with linguistic metaphors in news media or classroom talk. Without large-scale, systematically coded datasets, claims about intermodal reinforcement or complementarity remain largely speculative. This calls for further methodological development and interdisciplinary collaboration between corpus linguists, cognitive linguists, and semioticians.

Furthermore, while Conceptual Metaphor Theory (CMT) has been extended into multimodal domains (Forceville, 2009; Pérez-Sobrino et al., 2021), its application to real-world data across modalities remains uneven. Detailed comparative analysis of metaphorical mappings across verbal, visual, and gestural modes in varied corpora—such as political speeches, educational videos, or digital advertisements—would provide the kind of empirical grounding necessary to refine and test multimodal metaphor theory more robustly.

Additionally, socio-cultural context remains a critical but under-explored dimension in multimodal research. How specific modes are mobilized for meaning-making depends not only on cognitive structures, but also on genre conventions, institutional norms, and communicative technologies. These variables need to be more systematically integrated into

multimodal analysis, particularly through longitudinal or comparative studies of communicative practices across settings and cultures.

In sum, expanding the scope of textual analysis through a multimodal lens holds considerable promise, but demands more detailed, corpus-informed research to become fully operational. Only through such empirical deepening can multimodal theory provide not just a richer descriptive account, but also a predictive and explanatory framework that is grounded in actual language use. As multimodal corpora and annotation tools continue to develop, they will offer the means to explore these dimensions with the granularity and methodological rigor that the field currently lacks.

The Rubin's Vase image offers a particularly apt metaphor for understanding the complexities of multimodal meaning-making. In this classic Gestalt figure-ground illusion, viewers alternately perceive either a central vase or two opposing human profiles, depending on which elements are visually foregrounded. This visual ambiguity does not arise from any alteration in the image itself but from a shift in perceptual focus. The same visual stimulus gives rise to multiple, mutually exclusive interpretations based solely on contextual emphasis. In this way, Rubin's Vase becomes more than a psychological curiosity—it serves as a model for how meaning in communication is co-constructed through the relational dynamics between semiotic modes.

Multimodal discourse operates under similar perceptual principles. Language, image, gesture, and other modes rarely function in isolation; rather, they shape and are shaped by one another. The meaning of a text is not simply inscribed in its verbal content, but emerges through the dynamic interplay of all semiotic resources present. As Kress and van Leeuwen (2006) argue, modes operate in concert, each bringing distinct affordances to the act of communication. The Rubin's Vase metaphor captures this interaction with striking clarity: just as the viewer cannot perceive the vase and the faces simultaneously, attention in multimodal communication is always selective, context-dependent, and shaped by cultural and cognitive frames.

In textual analysis, this implies that meaning is not fixed or static but oscillatory and relational. A visual element may serve as the "ground" against which a verbal message stands out as the "figure," or vice versa, depending on how the text is framed, read, or situated. This challenges logocentric assumptions that meaning originates solely from linguistic content and invites analysts to consider how various semiotic elements modulate, reinforce, or even contradict one another. The figure-ground reversal seen in Rubin's Vase thus parallels how different modes can shift in salience during meaning construction, revealing the cognitive and perceptual fluidity inherent in multimodal texts.

Moreover, the image metaphorically underscores a broader theoretical implication: that understanding is never absolute but is mediated by

perspective and orientation. Just as the viewer must constantly negotiate between seeing the vase or the profiles, so too must readers and analysts navigate the fluctuating prominence of different semiotic components. This reinforces the importance of training critical attention toward how modes interact, not only in terms of presence but also in terms of prominence, cohesion, and framing within the communicative act (Bateman, 2014; Bezemer & Kress, 2016).

In pedagogical contexts, this metaphor proves particularly useful. When teaching students to analyze multimodal texts—be they political speeches, advertisements, or digital media—the Rubin's Vase analogy clarifies that meaning is not "in" any single mode. Instead, it invites learners to ask what is being foregrounded, what remains in the background, and how these perceptual choices shape interpretation. In this sense, the Rubin's Vase not only serves as a model for theoretical reflection but also as a pedagogical device to cultivate multimodal awareness.

Thus, employing Rubin's Vase as a metaphor in the study of multimodality illuminates the shifting, interdependent, and perspectival nature of meaning. It reveals that multimodal analysis, like the image itself, requires an attentiveness to perceptual oscillation—an ability to hold multiple semiotic possibilities in view and to trace the cognitive movement between them.

The Rubin's Vase remains a paradigmatic illustration of figure-ground perception, first introduced by Danish psychologist Edgar Rubin in 1915. His seminal work on visual perception emphasized the human ability to distinguish a central object (the figure) from its surrounding context (the ground), revealing that these roles are not fixed but can reverse depending on the observer's cognitive orientation (Rubin, 1915). This phenomenon encapsulates a foundational insight into human perceptual organization: that the boundaries between figure and ground are perceptually malleable, shaped by attention, context, and interpretive intent.

Rubin's insight was taken up and developed by early Gestalt psychologists, including Max Wertheimer, Wolfgang Köhler, and Kurt Koffka, who argued that perception operates according to holistic principles. These scholars contended that individuals do not perceive isolated sensory inputs but structured wholes, in which the relation between parts contributes fundamentally to meaning (Wertheimer, 1923; Köhler, 1947; Koffka, 1935). The Rubin's Vase illusion thus became emblematic of the Gestalt principle that perception is not merely a matter of sensory input, but a dynamic organization that prioritizes coherence, simplicity, and closure.

Later cognitive research has continued to investigate the Rubin's Vase as a model for perceptual reversibility, examining how neural mechanisms facilitate the alternating interpretation of a single stimulus. Studies in

neuroscience, such as those by Hesselmann, Kell, Eger, and Kleinschmidt (2008), have demonstrated that shifts in perception between figure and ground are predicted by prestimulus brain activity, particularly in areas associated with attentional modulation. Their findings suggest that the perceptual system is not passively driven by stimuli, but is actively shaped by top-down processes that influence what emerges as salient or backgrounded.

Beyond visual perception, the theoretical implications of Rubin's Vase have extended into cognitive linguistics and semiotics. Scholars such as Leonard Talmy (2000) and Ronald Langacker (2008) have applied the figure-ground schema to language, analysing how speakers linguistically frame events by designating certain elements as prominent (figure) and others as supportive context (ground). This linguistic application reinforces the idea that meaning is not absolute, but depends on patterns of salience and backgrounding—concepts that resonate strongly with multimodal analysis. In texts that integrate verbal, visual, and gestural modes, what is treated as the "figure" can shift depending on design choices, cultural expectations, or the interpretive strategies of the viewer.

Thus, the Rubin's Vase functions not only as a striking optical illusion but also as a theoretical metaphor for meaning construction across modalities. It demonstrates that interpretation is both perspectival and relational, shaped by perceptual focus and systemic context. In multimodal communication, as in Gestalt perception, modes co-exist in a fluid hierarchy, with salience and meaning arising from their interaction. This perspective supports the view that multimodal analysis must attend not only to the presence of different modes but to their orchestration, framing, and interplay—what Kress and van Leeuwen (2006) have called the "grammar of visual design."

CONCLUSIONS

Multimodality is not merely an additional layer to language; it is integral to the very structure of communication and essential for constructing and maintaining mental representations of reality. In the context of English, the combined use of semiotic modes—spoken and written language, gesture, images, sound, and spatial elements—forms a complex web through which we negotiate, share, and adapt our conceptual understandings of the world. These multimodal interactions are foundational to how we convey meaning, think about abstract concepts, and engage with others within specific cultural and social contexts.

Through the lens of multimodal analysis, English communication emerges as more than just a system of linguistic signs; it becomes a dynamic cognitive ecology where meaning is not isolated within the boundaries of language itself but is co-constructed through a range of interconnected modes. The significance of this approach goes beyond its ability to analyze how people communicate in real-world settings; it

extends to how human cognition processes and organizes knowledge. As users of English, we rely on various modes to reinforce and complement one another, creating a cohesive framework that supports how we understand and express thoughts, ideas, and emotions.

Nowhere is this clearer than in the study of grammar. Far from being a static system of syntactic and morphological rules, grammar emerges as a multimodal phenomenon—embodied, distributed, and context-dependent. Multimodal Metaphor Theory and Cognitive Grammar have shown that grammatical structures are often underpinned by embodied schemas and conceptual mappings, which manifest across gesture, prosody, spatial positioning, and visual representation. Theoretical grammar, when expanded through a multimodal lens, allows scholars to explore how meaning is dynamically co-constructed not just in spoken and written forms, but through gestures that align with syntactic boundaries, prosodic contours that frame information structure, and even visual cues that index grammatical relations. This reorientation challenges monomodal traditions and opens the field to a more ecologically valid understanding of how grammar functions in everyday communication.

Likewise, multimodality plays a crucial role in political discourse, where the strategic orchestration of verbal and non-verbal modes serves both rhetorical and ideological purposes. Political communication today operates across media platforms in ways that rely heavily on visual imagery, voice modulation, bodily posture, and spatial arrangement—each of which interacts with linguistic choices to shape public perception. Multimodal analysis uncovers how gestures reinforce ideological stances, how visual metaphors sustain national or moral narratives, and how spatial configurations in debates or public appearances contribute to the projection of authority or vulnerability. These insights reveal the limitations of purely textual or linguistic analyses of political speech and underscore the need for integrated frameworks that account for the full semiotic repertoire available to speakers and institutions alike.

This multimodal perspective is especially powerful in academic and practical fields such as applied linguistics, philology, translation studies, and discourse analysis, where a richer understanding of meaning-making can inform teaching, research, and professional practice. In English philology, for instance, recognizing the interdependence of verbal and non-verbal resources challenges traditional logocentric views that prioritize language alone and encourages a more comprehensive examination of how texts, genres, and communication systems operate. By integrating multimodal principles, scholars and educators can develop new methodologies that account for the multiplicity of semiotic resources at play in both historical and contemporary English texts.

Moreover, in applied philology and pedagogy, multimodal frameworks hold significant implications for teaching practices, particularly in how English is taught to non-native speakers. Instruction that emphasizes not

just the grammatical structures of the language but also incorporates visual, auditory, and gestural elements fosters a more holistic understanding of communication. Whether through the teaching of metaphors, the analysis of visual culture in texts, or the application of multimodal translation strategies, this approach better prepares learners and professionals to navigate the complexity of meaning-making in real-world scenarios.

Ultimately, multimodality broadens the scope of English studies, opening new avenues for research that engage with the full spectrum of human communication. As both a theoretical paradigm and a practical framework, multimodality enriches our understanding of English not just as a system of signs but as a dynamic, cognitive environment that mirrors the full diversity of human experience. It underscores the interconnectedness of language, thought, and culture, offering fresh insights into the ways in which communication shapes and is shaped by the world around us. By incorporating this perspective into research and practice—from the study of grammar to political discourse—scholars and practitioners can contribute to a deeper, more nuanced understanding of language in the contemporary world.

The integration of multimodal approaches into English studies also responds to an epistemological shift in the humanities and social sciences, where language is increasingly recognized not as an isolated system of code, but as embedded within a broader semiotic landscape. This perspective aligns with systemic functional linguistics (SFL), particularly the work of Halliday and Matthiessen (2014), which conceives grammar as a resource for making meaning within social contexts rather than a mere set of syntactic rules. Multimodal theorists have extended this principle, suggesting that just as grammar expresses ideational, interpersonal, and textual meanings in verbal language, so too do images, gestures, and spatial arrangements perform analogous functions in other semiotic systems (Kress & van Leeuwen, 2001).

Evidence from gesture studies supports this claim. McNeill (2005) and Kendon (2004) have shown that gestures not only accompany speech but often perform syntactic, referential, and discourse-organizing roles. For instance, pointing gestures can function deictically, substituting or complementing locative expressions in spoken discourse, while iconic gestures frequently anticipate or scaffold complex syntactic constructions. This co-articulation between modes exemplifies how grammar, when viewed through a multimodal lens, emerges as distributed across bodily action, visual representation, and spoken language.

The field of discourse analysis, particularly critical discourse analysis (CDA), also benefits from this multimodal orientation. Scholars like Machin and Mayr (2012) argue that multimodality is essential for understanding how ideology is embedded and naturalized in communicative practices. In political discourse, for example, the

deployment of multimodal metaphor often serves persuasive or hegemonic functions. Forceville (2009) identifies visual metaphors in political cartoons and media imagery that, through analogy and juxtaposition, convey complex ideological meanings more potently than language alone could. These metaphors are not decorative but form part of the conceptual architecture through which political positions are legitimized and contested.

Consider, for example, political campaign videos or televised debates. Research shows that politicians use carefully choreographed gestures, gaze direction, posture, and voice modulation to construct specific identities—such as competence, empathy, or strength—which complement their verbal strategies (Poggi, D'Errico, & Vincze, 2010). The multimodal orchestration of these performances functions as a kind of "grammaticality" of public discourse, where congruence between verbal and non-verbal modes enhances persuasiveness, and dissonance can result in perceived inauthenticity or rhetorical failure.

Further support for multimodality's impact on grammar and communication comes from studies in digital media. In computer-mediated environments, grammatical conventions are frequently reshaped by the interplay of text, image, emoji, layout, and hyperlinking. Jewitt et al. (2016) and Ledin and Machin (2016) have demonstrated how visual and spatial design choices in web interfaces and social media posts influence reading paths, emphasis, and syntactic interpretation. These findings indicate that digital grammar is increasingly multimodal, requiring users to interpret meaning from complex semiotic assemblages that go well beyond linear, alphabetic language.

Despite these advances, many theoretical grammar models remain rooted in formalist paradigms that treat multimodality as peripheral or even disruptive. As such, there is an urgent need for integrated theoretical frameworks—such as the proposed *Multimodal Grammar Theory* or *Embodied Multimodal Metaphor Theory*—which account for the cognitive, social, and semiotic dimensions of language in use. These frameworks would not only bridge traditional linguistic concerns with contemporary communication practices but would also better reflect how humans learn, process, and produce meaning across contexts and media.

Such a rethinking also holds implications for language pedagogy, especially in English as a second or foreign language contexts. As Kress (2010) and Cope and Kalantzis (2000) have argued, multimodal literacy is increasingly central to communicative competence. Teaching grammar through multimodal materials—such as video, gesture annotation, visual storytelling, or speech-gesture alignment—not only enhances learner engagement but also aligns instructional practices with how language functions in the real world. Moreover, by foregrounding metaphor and embodiment in grammar instruction, learners gain tools for conceptual flexibility, which is particularly important in cross-cultural and academic communication.

In sum, the multimodal turn in language studies compels a profound reconceptualization of grammar, discourse, and communicative practice. It urges scholars to move beyond narrow structuralist or logocentric models and to embrace a view of language as situated, embodied, and semiotically rich. Whether in the analysis of political rhetoric, the revision of theoretical grammar, or the design of inclusive language curricula, multimodality offers a robust framework for understanding how meaning is made, negotiated, and contested in contemporary society. As we advance into increasingly digital, visual, and performative forms of interaction, such a framework is not merely desirable but essential for maintaining the relevance and rigor of English language scholarship.

When we return to the domain of English philology and language education, the insights derived from Rubin's Vase and its associated figure-ground dynamics offer profound implications for how meaning is theorized and analyzed within linguistic systems. Philology, traditionally concerned with the historical, formal, and interpretive dimensions of texts, has often privileged the verbal code as the primary site of meaning. However, the multimodal turn in linguistics and discourse studies challenges this logocentric assumption, inviting scholars to consider how verbal elements are co-articulated with visual, spatial, and embodied modes in the full construction of meaning.

From this perspective, the figure-ground principle serves not only as a perceptual model but as a cognitive schema for understanding how communicative focus is established within language. In both spoken and written English, linguistic choices are not neutral but are shaped by the speaker or writer's framing of what is foregrounded as central and what is relegated to the background. This dynamic is evident in the use of grammatical constructions such as topicalization, clefting, voice, and information structure, all of which serve to guide attention and establish narrative or argumentative salience. The structural forms of English grammar can therefore be viewed as cognitive tools for managing figure-ground relationships across time and context.

Moreover, multimodal analysis reveals that figure-ground relations are not only realized grammatically but are often distributed across semiotic modes. For example, in a political speech, the verbal text may downplay or obscure certain ideologies, while visual staging, camera angles, and body posture foreground others. In digital discourse, the interplay of text, emoji, hyperlink placement, and graphic layout continuously reassigns figure-ground configurations to shape how meaning is accessed and interpreted. These examples illustrate that the philological object—what constitutes the "text"—has fundamentally changed. It is no longer confined to the printed page or verbal utterance but now encompasses a constellation of modes whose interaction generates significance.

In English language pedagogy, these theoretical insights translate into pedagogical imperatives. Teaching grammar, discourse, or literature

without acknowledging the multimodal and perceptually grounded nature of meaning risks misrepresenting how language actually works in situated communication. Instead, a multimodal approach to English instruction cultivates awareness of how learners navigate and produce meaning through a combination of verbal, visual, and embodied resources. It also prepares students to become more critical consumers and producers of texts in a media-saturated environment where interpretation is shaped as much by layout, design, and gesture as by syntactic rules.

Ultimately, returning to English philology through the lens of figure-ground theory and multimodal cognition reveals language as part of a larger meaning-making system—an ecology of interaction in which grammar, perception, and modality converge. Rubin's Vase, then, is more than a perceptual curiosity; it is a metaphor for the interpretive fluidity inherent in all communication. Just as viewers must toggle between two interpretations of the image, language users and analysts must remain alert to the shifting saliencies that structure meaning across modes. In this way, multimodal frameworks not only enrich philological analysis but redefine it, grounding the study of English in the full spectrum of human sense-making.

The integration of multimodal analysis into the study of English philology and language pedagogy represents a paradigm shift in the understanding of language and communication. Traditionally, philology has been concerned primarily with the historical, syntactic, and semantic aspects of texts, often focusing on the verbal mode as the central unit of meaning construction. However, this view is increasingly challenged by contemporary approaches that emphasize the interplay between language and other semiotic modes, such as gesture, image, and spatial positioning, in shaping meaning. Recent scholarship in multimodal linguistics and cognitive semiotics has demonstrated that meaning is not confined to verbal signs alone but is co-constructed through a complex network of interconnected modes that include visual, auditory, and bodily elements (Kress & van Leeuwen, 2006; Jewitt, 2017). This recognition of the multimodal nature of meaning-making has profound implications for both textual analysis and language instruction.

In the realm of textual analysis, contemporary linguists have begun to incorporate multimodal frameworks that allow for a richer understanding of how meaning is constructed in diverse communicative contexts. For example, Lei Yu and Yang Xu (2021) introduced the Syntactic Frame Extension Model (SFEM), which integrates multimodal knowledge to predict how new linguistic compositions emerge over time. This model highlights the importance of perceptual and conceptual grounding in the development of language, suggesting that the meaning of verbs, for instance, extends beyond purely linguistic frameworks to include the embodied and visual modes through which they are realized. Such research emphasizes that language does not operate in isolation but is constantly influenced by our sensory engagement with the world and our

cognitive processing of multimodal stimuli. This expanded view of linguistic meaning challenges the traditional focus on the syntactic structure of language, proposing instead that meaning arises from a dynamic interaction between linguistic signs and their contextual, multimodal environment.

Furthermore, the work of Michaela Mahlberg (2018) exemplifies how corpus linguistics, traditionally a tool for examining the frequency and patterns of words, can be adapted to explore the social and multimodal dimensions of language. Through the CLiC (Corpus Linguistics in Context) project, Mahlberg examines how literary texts engage with visual and social contexts, revealing how language use both reflects and shapes societal norms and perceptions. Her work underscores the role of language not just as a system of abstract signs but as a social practice that interacts with and responds to visual and cultural cues. The recognition that textual meaning can be influenced by external, non-verbal factors expands the scope of philological analysis, encouraging a more integrated approach that attends to the visual and social dimensions of texts.

In addition to these developments in corpus linguistics, research in psycholinguistics, particularly the work of Aslı Özyürek and her colleagues at the Max Planck Institute for Psycholinguistics, has further illuminated the multimodal nature of communication. Özyürek's research focuses on the interaction between gesture and speech, showing that these two modes work together to enhance comprehension and facilitate communication (Özyürek, 2014). Her studies reveal that gestures are not simply supplementary to speech but play an essential role in shaping how we understand and process linguistic meaning. This research aligns with the findings in cognitive linguistics, where the body and perceptual experience are considered central to the construction of meaning (Lakoff, 1987; Johnson, 2007). The interaction between verbal and non-verbal modes, as demonstrated by Özyürek's work, provides crucial insights for understanding how multimodal resources are used in everyday communication and how they contribute to the overall meaning-making process.

These contemporary studies in multimodal linguistics highlight the necessity of revisiting traditional approaches to philology, particularly in the teaching and analysis of English. As language scholars and educators, it is imperative to recognize that meaning in English texts—whether spoken, written, or visual—is not merely a product of syntactic and lexical choices but arises from a dynamic interaction between multiple semiotic resources. The integration of visual, auditory, and gestural elements with verbal language challenges the logocentric tradition that prioritizes the written or spoken word and encourages a more holistic understanding of how meaning is generated and interpreted.

In language pedagogy, these insights have profound implications for teaching English as a second or foreign language. By incorporating

multimodal resources such as gesture, visual aids, and digital media, educators can create more engaging and effective learning environments that ground grammatical concepts in perceptual and embodied experience. This approach aligns with the principles of multimodal literacy, which advocates for the development of students' ability to interpret and produce meaning across multiple modes (Kress, 2010; Jewitt et al., 2016). In doing so, it fosters not only linguistic competence but also critical thinking, creativity, and adaptability—skills that are essential for navigating the increasingly complex multimodal communication environments of the modern world.

Moreover, this shift towards multimodal analysis and pedagogy reflects a broader movement within the humanities to reconsider the boundaries of textuality. As Kress and van Leeuwen (2006) argue, the study of texts should no longer be confined to written or spoken language but must account for the visual, spatial, and embodied elements that are integral to meaning-making. This multimodal turn challenges scholars to rethink the very nature of texts and language, prompting a redefinition of the philological object to encompass the full range of semiotic resources that shape how we communicate and interpret the world.

In conclusion, the integration of multimodal frameworks into English philology and language instruction offers a more comprehensive and dynamic understanding of how meaning is constructed and communicated. The work of contemporary scholars such as Yu and Xu (2021), Mahlberg (2018), and Özyürek (2014) demonstrates the growing recognition that meaning-making is a multimodal process, shaped by the interaction between language and other semiotic modes. By adopting this perspective, philology and language pedagogy can move beyond the confines of traditional logocentrism, offering a more nuanced and holistic approach to the study and teaching of English. This shift not only enhances our understanding of grammar and language structure but also prepares learners to navigate the complexities of meaning-making in a multimodal world, ultimately shaping the future of language studies and pedagogy.

REFERENCES

Bateman, J. A. (2014). *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. Routledge.

Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, Research, and Analysis – A Problem-Oriented Introduction*. De Gruyter Mouton.

Bezemer, J., & Kress, G. (2016). *Multimodality, Learning, and Communication: A Social Semiotic Frame*. Routledge.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 483–501). Cambridge University Press.

Forceville, C. (2007). Multimodal metaphor in ten Dutch TV commercials. *Public Journal of Semiotics, 1*(1), 15–34.

Forceville, C., & Urios-Aparisi, E. (Eds.). (2009). *Multimodal Metaphor*. Mouton de Gruyter.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge.

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Rubin-vase illusion perception is predicted by prestimulus activity and connectivity in the visual cortex. *Journal of Cognitive Neuroscience, 20*(7), 1390–1401.

Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing Multimodality*. Routledge.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright.

Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge.

Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd ed.). Routledge.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.

Ledin, P., & Machin, D. (2016). *Doing Visual Analysis: From Theory to Practice*. SAGE Publications.

Mahlberg, M. (2018). *Corpus Linguistics and the Study of Transitivity in Literature: Insights from CLiC*. University of Birmingham Press.

McNeill, D. (2005). *Gesture and Thought*. University of Chicago Press.

Özyürek, A. (2014). The role of gesture in the comprehension of speech: How gestures and speech converge in communication. *Max Planck Institute for Psycholinguistics*.

Rubin, E. (1915). *Synsoplevede Figurer* [Visual Figures]. Gyldendalske

Boghandel.
Talmy, L. (2000). *Toward a Cognitive Semantics, Volume I: Concept Structuring Systems*. MIT Press.
van Leeuwen, T. (2005). *Introducing Social Semiotics*. Routledge.
van Leeuwen, T. (2006). Towards a semiotics of typography. *Information Design Journal + Document Design*, *14*(2), 139–155.
Wertheimer, M. (1923). Laws of organization in perceptual forms. *Psychologische Forschung*, *4*(1), 301–350.
Yu, L., & Xu, Y. (2021). Syntactic frame extension model: Multimodal knowledge and new linguistic compositions. *arXiv*. Retrieved from https://arxiv.org/abs/2109.04652
Wichmann, A. (2000). *Intonation in Text and Discourse: Beginnings, Middles, and Ends*. Longman.

Afterword

As this book draws to a close, it is worth pausing to reflect on the evolving intellectual landscape in which it was written. The intersections between cognition, language, and multimodality continue to unfold in new directions, offering fertile ground for inquiry, collaboration, and reinterpretation.

This work was never meant to offer final answers, but rather to open a space for dialogue—between disciplines, between modes of communication, and between readers and texts. It reaffirms a belief in the value of humanistic thinking, especially in times where rapid technological advancement challenges the depth and nuance of traditional scholarship.

In returning to the humanities, we return to what it means to understand, to empathize, and to interpret. Whether through a classical painting, a modern-day meme, or a political speech shaped by multimodal strategies, the goal remains the same: to read meaningfully and think deeply. Just as Rubens captured the essence of human emotion in movement and color, today we must learn to "read" across modalities with the same sensitivity and imagination.

If this book contributes, even in a small way, to advancing such a mindset—rooted in the belief that meaning is always richer than it first appears—then it will have served its purpose.

Let us continue to read between the lines, across the images, and through the silences—keeping the human at the center of our interpretive practices.

Information about the author:

A. Khodorenko

Background

The intersection of cognitive linguistics, multimodal discourse, and English philology. A foundation in philological studies, translation, and teaching. The exploration of how language, visual elements, and cultural context combine to shape meaning. An approach grounded in a humanistic tradition, emphasizing interpretive thinking, creativity, and a nuanced understanding of communication across various media.

Research Interests

Cognitive framing and multimodal metaphor. The semiotics of visual-verbal texts. The analysis of discourse in contemporary and digital forms of communication. An interest in the subtle ways language, thought, and art interact, and in helping others engage with the richness and depth of meaning embedded in multimodal texts.

Khodorenko, A. (2025). *Cognitive and multimodal perspectives in contemporary English philology: Theory, analysis, and applications*. Dnipro: Serednyak T.K. – 73 p. DOI 10.5281/zenodo.15403690.

ISBN 9-786178540579