

---

# Exploration of the Intersection of Corpus Linguistics and Language Science

Hakan Cangır  
*Ankara University, Turkey & Erfurt University, Germany*

Kutay Uzun  
*Trakya University, Turkey*

Taner Can  
*TED University, Turkey*

Enis Oğuz  
*Middle East Technical University, Turkey*

Vice President of Editorial  
Managing Editor of Acquisitions  
Managing Editor of Book Development  
Production Manager  
Cover Design

Melissa Wagner  
Mikaela Felty  
Jocelynn Hessler  
Mike Brehm  
Phillip Shickler

Published in the United States of America by

IGI Global Scientific Publishing  
701 East Chocolate Avenue  
Hershey, PA, 17033, USA  
Tel: 717-533-8845  
Fax: 717-533-8661  
Website: <https://www.igi-global.com> E-mail: [cust@igi-global.com](mailto:cust@igi-global.com)

Copyright © 2025 by IGI Global Scientific Publishing. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global Scientific Publishing of the trademark or registered trademark.

#### Library of Congress Cataloging-in-Publication Data

Names: Cangir, Hakan, editor. | Uzun, Kutay, editor. | Can, Taner, editor.  
| Oğuz, Enis, editor.

Title: Exploration of the intersection of corpus linguistics and language science / edited by Hakan Cangir, Kutay Uzun, Taner Can, Enis Oğuz.

Description: Hershey PA : IGI Global Scientific Publishing, 2025. |

Includes bibliographical references and index. | Summary: "This book aims to examine the dynamic interplay between corpus linguistics and various subfields within the broader domain of language sciences"--  
Provided by publisher.

Identifiers: LCCN 2024055697 (print) | LCCN 2024055698 (ebook) | ISBN 9798369381465 (hardcover) | ISBN 9798369381472 (paperback) | ISBN 9798369381489 (ebook)

Subjects: LCSH: Corpora (Linguistics) | LCGFT: Essays.

Classification: LCC P128.C68 E96 2025 (print) | LCC P128.C68 (ebook) |  
DDC 410.1/88--dc23/eng/20241230

LC record available at <https://lcn.loc.gov/2024055697>

LC ebook record available at <https://lcn.loc.gov/2024055698>

#### British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material.

The views expressed in this book are those of the authors, but not necessarily of the publisher.

This book contains information sourced from authentic and highly regarded references, with reasonable efforts made to ensure the reliability of the data and information presented. The authors, editors, and publisher believe the information in this book to be accurate and true as of the date of publication. Every effort has been made to trace and credit the copyright holders of all materials included. However, the authors, editors, and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. Should any copyright material be found unacknowledged, please inform the publisher so that corrections may be made in future reprints.

---

# Table of Contents

Preface..... xi

Acknowledgment..... xx

## Section 1 Foundations

### Chapter 1

Connecting Corpus Linguistics and Psycholinguistics: A Dual Approach to Collocation Research ..... 1  
*Irene Fioravanti, University for Foreigners of Perugia, Italy*

### Chapter 2

Corpus Linguistics Meets Psycholinguistics: Insights From Turkish-English Collocations ..... 27  
*Hakan Cangır, Ankara University, Turkey*

### Chapter 3

Why Go Digital? Literary Studies in the Age of Digitalisation..... 55  
*Taner Can, TED University, Turkey*

## Section 2 Corpus Linguistics in Language Learning and Acquisition

### Chapter 4

Morphological Complexity Development Across Proficiency Levels in English as a Foreign Language ..... 73  
*Kutay Uzun, Trakya University, Turkey*

### Chapter 5

L2 Preposition Errors in a Learner Corpus: Evidence of U-Shaped Development and L1 Influence.. 89  
*Enis Oğuz, Middle East Technical University, Turkey*

### Chapter 6

The Use of Prepositional Phrase-Based Bundles in Essays Written by Turkish Learners of English 109  
*Ömer Faruk Kaya, Bursa Technical University, Turkey*

### Chapter 7

Using a Speech Corpus to Raise Phonemic Awareness Among L2 English Speakers: Production and Substitution of Dental Fricatives..... 133  
*Ceren Yağmur Öztürk, TED University, Turkey*

---

<b>Chapter 8</b>	
Learning English From Videogame Characters: Results From Corpus Data.....	157
<i>Francesca Bianchi, University of Salento, Italy</i>	

**Section 3**  
**Corpus-Based/Corpus-Assisted Linguistic Analysis**

<b>Chapter 9</b>	
Investigating the Lexical Associations in the L3 Italian Mental Lexicon .....	191
<i>Cumhur Kuzu, Ankara University, Turkey</i>	
<i>Hakan Cangır, Ankara University, Turkey</i>	

<b>Chapter 10</b>	
It's High Time: A Corpus and NLP-Based Investigation of the Time Metaphors in Turkish .....	223
<i>Alper Kumcu, Hacettepe University, Turkey</i>	

<b>Chapter 11</b>	
Cognitive Patterns in Learner Texts: Exploring Semantic and Semiotic Dimensions Through Topic Modeling.....	253
<i>Fatih Ünal Bozdağ, Osmaniye Korkut Ata University, Turkey</i>	

<b>Chapter 12</b>	
Gender-Sensitive Language in Foreign Language Education: An Analysis of Textbooks for Italian as a Foreign Language .....	279
<i>Chiara Dabbene, University of Milan, Italy</i>	

<b>Chapter 13</b>	
The Champions of If-Constructions in Literary Studies: A Comparative Analysis of If- Constructions in L2 Learners' M.A. Theses and Ph.D Dissertations .....	307
<i>Özkan Kırmızı, Karabuk University, Turkey</i>	
<i>Haticetül Kübra Er, Erzurum Technical University, Turkey</i>	

<b>Chapter 14</b>	
Corpus Linguistics and Literary Translation: A Contribution to Consistency .....	327
<i>Maria João Ferro, Centre for English Studies, University of Lisbon, Portugal</i>	

<b>Compilation of References</b> .....	357
<b>About the Contributors</b> .....	409
<b>Index</b> .....	413

---

# Chapter 8

## Learning English From Videogame Characters: Results From Corpus Data

Francesca Bianchi

 <https://orcid.org/0000-0002-9382-4521>

University of Salento, Italy

### ABSTRACT

*Studies have shown that non-anglophone students greatly benefit from watching pop-cultural products such as films and videogames in English, as these support unsupervised language learning, and that scripted dialogue in films and in videogames both resemble spontaneous speech in many ways. But are there any lexico-grammatical features that a person is more likely to encounter, and consequently learn, when playing videogames compared to when watching films? To answer this question this study contrasts scripted dialogue in videogames to film dialogue and spontaneous interactions, at the level of part-of-speech categories. Results highlight a small set of features that are significantly more frequent in videogame scripted dialogue. Such features are mostly connected to the ludic aims of videogames.*

### INTRODUCTION

Youngsters in non-English speaking countries have frequent informal contact with English when engaging with multimedia products (e.g., Sundqvist, 2009; Verspoor et al., 2011). Multimedia products is an umbrella term encompassing a wide range of works where the aural and the visual channels combine to create the final object. These include widespread multimedia pop-cultural artefacts, such as films, TV series, and videogames. Exposure to all types of multimedia has beneficial effects on a learner's language competence in the foreign language in general and as regards vocabulary acquisition in particular (e.g., Caruana, 2003, 2006 for TV input and Sundqvist, 2009 for game playing, among many). But do these different types of products expose learners to the same linguistic events, quantitatively and/or qualitatively?

Telecinematic dialogue, such as the characters' lines in films and TV series, and dialogue between characters of a videogame are forms of scripted text (or performed text) meant to resemble real-life discourse, but produced for asynchronous fruition by an extradiegetic audience (the viewer or the player). This means that they can both be expected to be slightly more formal than unscripted dialogue (Werner,

DOI: 10.4018/979-8-3693-8146-5.ch008

Copyright ©2025, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishing is prohibited.

---

2021). Indeed, although performed text often simulates dialogue and is used to convey (pseudo-)intimacy between stage participants, scripted language has been preplanned and carefully edited for specific purposes and to achieve a particular effect. The result is a form of language that lacks those conversational features which are directly connected to real-time processing constraints. Furthermore, the presence of an extradiegetic audience requires performed text to be more explicit than spontaneous conversation (see Werner, 2022, p. 8-9).

However, some degree of difference between these two forms of scripted dialogue is also foreseeable, especially if we consider that research on telecinematic discourse – i.e. films, TV series and sit-coms – has proven that non-linguistic factors, such as target medium, target broadcast time span, budget and production process, affect the language of performed dialogue (Bednarek, 2018).<sup>1</sup> Videogames are produced under totally different circumstances and with ludic aims which are completely different from those of telecinematic products. In particular, in videogames, non-player character (NPC) talk serves the diegetic functions of describing the fictional world of the game and supporting the progression of the narrative, alongside the ludic function of giving instructions to the player (Domsch, 2017), a function that is entirely absent in telecinematic products.

Spontaneous real-life conversation and telecinematic dialogue have been extensively analyzed by linguists, both independently and in comparison with one another, providing substantial evidence of the lexical and grammatical features they contribute to language learning. On the other hand, studies on the scripted dialogue between characters in videogames are still scarce, most research on videogames having been carried out in fields other than linguistics (e.g. Fernández-Vara, 2014 on player-game relationship and the socio-cultural role of videogames; Bogost, 2007 and Paul, 2012 on the rhetoric of videogames; Fizek, 2012 on the study of videogame characters; Gee, 2014 on videogames as forms of communication), on the visual features of videogames (e.g. Stamenković & Jačević, 2019; Stamenković, 2023), on live dialogue between players (Ensslin, 2012; Ensslin & Finnegan, 2019; Graham & Dutt, 2019; Kiourti, 2019; Iaia, 2021; Rudge, 2019) or on game paratext (Ensslin, 2012, and more recently Álvarez-Bolado Sánchez & Álvarez de Mon, 2019 on videogame reviews; Balteiro, 2019 on forums; Gledhill, 2019 on tutorials and walkthroughs; Campos-Pardillos, 2019 on end-user agreements; Hancock, 2019 on manuals). To the best of our knowledge, no prior research has directly compared videogame dialogue with telecinematic dialogue. This study seeks to address this gap by examining and contrasting scripted dialogue in videogames with that in films. This will contribute to the broader aim of establishing the role that different forms of multimedia products can play in the unsupervised acquisition of L2 English.

The next section summarises the most relevant studies investigating variation across spoken registers, with a special focus on real-life conversation, telecinematic dialogue and scripted dialogue in videogames, along with the potential role of videogames in language acquisition. Subsequently, the paper will present the study's research questions, describe the materials and methods used and illustrate the findings. Finally, a concluding section will discuss the relevance of the findings from the perspective of language learning and suggest further research paths.

## REVIEW OF BACKGROUND LITERATURE

### Real-Life Dialogue

With reference to the English language, corpus-linguistic studies have investigated the specific features of a wide range of spoken registers. A fundamental role has been played by a method called ‘multidimensional analysis’.<sup>2</sup> Studies in this area have shown that face-to-face conversation is primarily a form of involved production (Biber, 1988) and that the context of situation – namely, place, participants and purposes – influences the frequency of certain grammatical features in a conversation (Biber & Conrad, 2009).

The features that characterize face-to-face conversation, compared to other types of registers, include frequent occurrences of questions generally, *wh*-questions, *wh*-clauses, *that*-clauses, *if*-clauses, other subordinators, final prepositions, the pro-verb *do*, contractions, first and second person pronouns, the pronoun *it*, general hedges, general emphatics, specific emphatics, demonstrative *be*, present tense, and infinitives. (Biber, 1988). General conversation also shows highly frequent use of lexical verbs (generally expressing stance), modal verbs, semi-modals, and adverbs (including locative adverbs), besides easily observable features, such as omissions of structural elements, attention getters, vocatives, and minimal responses (Biber & Conrad, 2009, Ch. 4). At the same time, spoken registers feature relatively few nouns and prepositions (Biber, 1988) and rather simple forms of nominal complexity, with few elements modifying the head noun, and rarely including both pre- and post-modification (Biber et al., 2021).

Syntax heavily relies on the use of dependent or embedded clauses (Biber et al. 2021), particularly the following types: *because*- and *if*-clauses; *that* complement clauses controlled by a verb (e.g. *I think [that] we better wait*); *wh* complement clauses controlled by a verb (e.g. *we didn’t know [what we were gonna be doing]*) (Biber & Conrad, 2009, Ch. 4).

The role of the context of situation is particularly evident in certain ‘specialised’ types of live interaction. For example, in conversations between students and faculty members during university office hours, deictics acquire specific quantitative relevance due to the frequent need to refer to written information. *If*-clauses and *because*-clauses are more frequent than in general conversation, as the goal is to solve problems collaboratively. Additionally, the verbs *want* and *need* are especially common. On the other hand, service encounters show a significantly higher use of the modal *can* followed by the pronoun *I*, as well as formulaic expressions (Biber & Conrad, 2009, Ch. 4).

### Telecinematic Dialogue

Dialogue in English-language films has received extensive attention from a wide range of scholars. An organic summary of the most relevant corpus-linguistic studies is offered by Pavesi (2022). Overall, despite its scripted nature, film dialogue compares to spontaneous interaction in several ways. Like the latter, it is characterized by features of *involved production*, including: personal pronouns, present tense markers, imperatives, private verbs, and qualifiers (Forchini, 2012); interjections, intensifiers, and colloquial questions (Zago, 2016); demonstratives (Pavesi, 2013).<sup>3</sup> As is the case with spontaneous dialogue, dependent or embedded clauses are frequent, including *wh*-clauses (Veirano Pinto, 2014; Zago, 2016), *to*-clauses controlled by verbs of desire, intention and decision (Veirano Pinto, 2014; Zago, 2016), and *that*-clauses controlled by a verb (Veirano Pinto, 2014); furthermore, according to Pavesi (2016), the percentages, patterns, and functions of demonstrative cleft clauses in film dialogue closely mirror those

in spoken English. Film dialogue also resembles spontaneous dialogue in terms of the variety of question patterns it presents (Ghia, 2014) and captures the geographic and socio-linguistic variations of English (Formentelli, 2014). At the same time, compared to spontaneous conversations, film interactions exhibit a significantly higher frequency of certain features. These include the demonstrative *this* (Pavesi, 2020), typically used to foreground facts and objects relevant to the narrative or to illustrate the interlocutors' personal spheres and perspectives (Pavesi, 2020); interrogatives (Ghia, 2014), used as implicit narrative techniques or effective ways to uncover details about characters and the story (Ghia, 2014), while also serving to provoke conflict and assert the questioner's authority (Ghia, 2014); and vocatives (Formentelli, 2014), which are used to strengthen the narrative and reveal characters' identities while affirming or redefining the relationships between them (Formentelli, 2014). Finally, a high degree of formulaic expressions can be observed in film dialogue compared to natural conversation, expressing politeness and emotionality (Freddi, 2011; Zago, 2021), alongside a higher frequency of pre- and post-modified noun phrases (Galiano, 2024).

Sit-coms and TV series have also been shown to successfully replicate the features of spontaneous conversation in both small-scale and large-scale investigations (Quaglio, 2008, 2009a, 2009b on the sit-com *Friends*; Bednarek, 2018 on the Sydney Corpus of Television Dialogue, a corpus comprising dialogue from 66 TV series). In particular, like face-to-face conversation, they show a statistically significant frequency of specific grammatical elements, such as first- and second-person pronouns and determiners, time/place adverbs *here* and *now*, intensifiers (e.g. *really*, *totally*, *very*), verbs of emotion and cognition (e.g. *love*, *likes*, *hated*, *believe*, *think*, *guess*, *want*), evaluative adjectives (e.g. *awesome*, *awful*, *funny*, *gorgeous*, *hilarious*, *horrible*), discourse markers (e.g. *well*), politeness markers (e.g. *please*, *thanks*), response and agreement/disagreement markers (e.g. *mm-mmm*, *mm hmmm*, *yeah*, *yep*, *yes*), interjections (e.g. *ah*, *gosh*, *oh*, *ugh*, *yay*, *yo*), hesitation markers (e.g. *uh*), and also of some lexical elements, such as informal words (e.g. *babe*, *fun*, *stuff*, *guys*), terms expressing kinship and relationship (e.g. *mom*, *brother*, *dad*, *friend*), and swear/taboo words (e.g. *ass*, *bitch*, *bullshit*, *damn*, *fucking*) (Bednarek, 2018). However, TV series differ from spontaneous conversation in many ways. On the one hand, there is overuse, compared to unscripted spoken dialogue, of interjections (Bednarek, 2012; Quaglio, 2008), routine formulae (e.g. *let's go*, *excuse me*, *thank you*, *I'm so sorry*), first- and second-person pronouns/determiners, terms of address, greetings, introductions, leave-takings, spatial and temporal deictics referring to the here and now, agreement responses (e.g. *alright*; *fine*; *of course*), contractions, swear/taboo words, slang or informal expressions, titles, evaluative adjectives (e.g. *best*, *crazy*, *wrong*), emotion lexis (e.g. *hurt*; *love*; *trust*; *worry*), imperatives, intensifiers, and word forms or phrases used to refer to aspects of the plot (Bednarek, 2018, Tab. 7.3, p. 130). Some of these features – in particular emphatic forms of expression, including adverbial intensifiers (e.g., *really*), some inserts (e.g., *wow*), stance markers (e.g., *of course*), and expletives/taboo terms (e.g., *damn*) – probably serve the purpose of keeping the audience engaged (Quaglio, 2008, 2009a, 2009b). On the other hand, at least some TV series have been found to have fewer instances of vague language (hedges, coordination tags, nouns of vague reference, and the discourse marker *you know*) than real-life conversations (Quaglio, 2008, 2009a, 2009b on *Friends*). Finally, an ample presence of questions (e.g. *can you*, *have you*, *what about*, *what do you*, *how did you*, *what are you going to*) and of phrases referring to the act of speaking itself (e.g. *to talk/tell*, *talk/ing about*, *talk to*; *are you talking about*; *tell her/him/me/them/you*, *told me/you*, *I told you*; *going to ask/call*) have been observed in TV series (Bednarek, 2018). Nevertheless, the language in TV series tends to resemble spontaneous conversation in terms of overarching patterns of grammatical

---

complexity, although it exhibits greater frequency of non-finite complement clauses and of nominal premodifiers (Formentelli & Zago, 2024).

## Videogame Scripted Dialogue

With the exception of multiplayer online role-playing games (MMORPGs) – which involve real-time interaction among multiple players and incorporate live written or spoken chats alongside gameplay, thus providing a different and wider variety of opportunities to enhance linguistic skills and support vocabulary development (e.g. Rankin et al., 2008; Peterson, 2010, 2011, 2012) – a primary source of linguistic input in videogames is in-game text and in particular NPC dialogue. However, in-game scripted language has only recently started to attract the attention of linguists, and some of the existing publications analyse both written and oral text jointly considered (e.g. Rodgers & Heidt, 2021; Heidt et al., 2023). The following paragraphs summarise the few existing linguistic studies on NPC talk.

Dixon (2022) analysed the scripted text of NPC dialogue in a corpus including four highly popular titles, for a total of 4.8 million words.<sup>4</sup> He distinguished the dialogue between NPCs, which he calls Immersive Speech, from the dialogue between NPCs and the player character/avatar, which he calls Interactive Speech.<sup>5</sup> The two registers were compared to real-life discourse using Biber's (1988) Involved versus Informational Production dimension (Dimension 1). Both Immersive Speech and Interactive Speech resulted towards the positive end of the involved-informational continuum (Dixon, 2022), though with mean dimension scores (23.61 and 29.8, respectively) that are slightly lower than that of face-to-face conversation (35.3). The author interpreted these results by stating that the spoken dialogue in these games exhibits a strong linguistic resemblance to real-world conversation, particularly when contrasted with other written registers found in the games (p. 166) and that interactive speech in the games more closely mirrors real-world speech than immersive speech (p. 166). However, the data also suggest the existence of linguistic similarities between the two spoken game registers and personal letters, as the latter scored just below the game registers at 19.5 along the *Involved versus Informational Production* dimension (p. 167). Thus, Dixon's results confirm the general expectation that scripted/performed dialogue tends to be slightly more formal than unscripted dialogue despite its attempt at emulating the latter (Werner, 2021).

Dixon (2022) also compared the spoken and written registers of the videogames to each other by applying multidimensional analysis methods and established four dimensions of variation specific to gaming discourse: Interactive Discourse vs. Information on Specific Entities; Interpretations of Past Events; Concise Technical Descriptions; and Procedures and Directives. Each dimension identifies clusters of co-occurring linguistic features that are valuable for distinguishing between different registers in videogames. In his corpus, both Immersive Speech and Interactive Speech sit at the positive end of the Interactive Discourse vs. Information on Specific Entities cline, which is indicative of highly involved communication among NPCs and between NPCs and the player avatar. The positive features include: present tense, contractions, the 100 most frequent verbs in the Longman spoken sub-corpus, modal verbs, first and second person pronouns, adverbs, verb *be*, and pronoun *it*; negative features include: prepositions, definite articles and proper nouns. The features at the negative end of the cline give clear and specific information about entities that are central to the gameplay, thus assisting the player in making decisions. This explains why dialogue in videogames is marked by a greater level of specificity than what is usually found in real-world conversation (p. 96) and why Interactive Speech showed slightly lower frequency of the co-occurring positive features of this dimension and more frequent use of the negative features, compared to Immersive Speech. Furthermore, in addition to indicating involvement,

---

features at the positive end of the cline are essential in videogame dialogue for providing direction to the player, warning them about potentially dangerous enemies, and describing probable future circumstances. Specifically, activity verbs are used to disclose the actions that must be performed in the game, as well as the intentions or capabilities of the individuals involved in those actions (p. 155).

No study has directly compared NPC talk and film dialogue. However, lexical coverage – defined as the vocabulary size required for readers to understand a text without external support – has been investigated in two separate studies, one examining in-game character talk (Bianchi, 2024) and the other analyzing film dialogue (Webb & Rodgers, 2009). Both studies employed the same analytical techniques and tools, which ensures the comparability of their results. A comparison of their results suggests that performed text in videogames exhibits a level of lexical difficulty comparable to that of British and American films. In fact, when it comes to NPC lines (Bianchi, 2024), a player should generally be familiar with the most frequent 3,000 word families to be able to guess unknown words in context (with internal variation ranging from 2,000 to 4,000 word families) and with the first 7,000 to fully comprehend the dialogue (with internal variation ranging from 5,000 to 10,000 word families). Similarly, in films (Webb & Rodgers, 2009), 3,000 word families are generally required to guess unknown words in context (with internal variation ranging from 3,000 to 4,000 word families), while 6,000 are on average necessary to fully understand the actors' lines (with internal variation ranging from 5,000 to 10,000 word families).

Finally, qualitative analysis of the scripted lines of non-English NPCs speaking in English in a limited set of videogames showed a good degree of similarity with exiting ELF (English as a Foreign Language) variations as described in the literature (Iaia, 2016, Chapter 7).

## **Videogames and Language Learning**

Videogames are largely recognized as “motivating, authentic, cognitively and linguistically complex, and effective learning environments” (Reinhart & Thorne 2020, p. 411). In particular, they offer an ideal environment for situated learning (Gee, 2006, 2008), where embodied experience is simulated (Gee, 2008) and players receive continuous and instantaneous feedback through trial-and-error, scaffolded interactions (Reinhardt, 2019). The general potential of videogames in language learning has been largely confirmed by empirical studies. Positive correlation has been found between frequency and length of engagement with videogames and language skills in the foreign language, particularly in terms of vocabulary knowledge (e.g. De Wilde et al. 2019; De Wilde & Eyckmans 2017; Kuppens 2010; Sylvén & Sundqvist 2012; Sundqvist & Wikström 2015; Rød & Calafato 2024, in Calafato & Clausen 2024; see also Poole, 2020, for a meta-analysis evidencing that game conditions are conducive to larger vocabulary gains more than non-game conditions).<sup>6</sup>

However, certain gameplay features are believed to be more effective than others in supporting incidental language learning. One of these is the presence of a narrative structure, which in videogames is accompanied by thematically coherent visual and linguistic contextualization cues (Domsch, 2017). In particular, research indicates that vocabulary is more effectively acquired and retained over time when it is presented in narrative contexts rather than in non-narrative ones (Reinhardt, 2019). On the other hand, time constraints on the player and speed of gameplay are expected to negatively impact on language acquisition. Specific game genres based on narratives and rich in linguistic input, such as adventure and interactive fiction, have been suggested to be optimal for L2 learning, especially when they enable self-paced progress, include features like captions and repeatability, and avoid penalizing players for taking their time (Reinhart & Thorne 2020, p. 417). Once again, empirical studies have provided confirmation

---

of this. For example, driving games, which include few dialogic situations and focus on controlling a vehicle while navigating obstacles, have been found to negatively correlate with vocabulary learning (Calafato & Clausen, 2024).

In single-player games with narrative structure, like the ones considered in this study, NPC dialogue is a primary source of linguistic input for L2 players. Input is an initial, fundamental element for second language acquisition as it provides positive evidence about what is possible and used in a given language. However, for spontaneous language acquisition to take place, exposure to linguistic input is to be accompanied by other factors. One particularly relevant factor is noticing (Schmidt, 1990, 1995, 2001, 2010; Schmidt & Frota, 1986). Noticing refers to the cognitive process of consciously registering language input: only the lexical or grammatical structures in the input which are noticed by the learner will be integrated into the interlanguage system (Schmidt 1990). Noticing is influenced by many factors, including frequency (Ellis, 2002; Gass & Mackey, 2002), salience (Schmidt & Frota, 1986), and the usefulness of language for performing a task or achieving a goal that is meaningful to the learner (e.g. Loschky & Bley-Vroman 1993; Newton, 1995). Language processing is highly sensitive to usage frequency at all levels of language representation, including vocabulary and grammar. In particular, it has been observed that both type (word) and token (word form) frequency play a crucial role in learning productive constructions (Bybee, 2008). Conventionalized sequences do not necessarily need to be highly frequent, as a few repetitions are sufficient to establish a sequence as conventional for native speakers (Bybee, 2008). The notion of frequency is particularly relevant in the current paper, as it is a textual feature that can be easily measured using corpus linguistic tools. Corpus linguistics can tell us what lexico-grammatical features are particularly frequent in videogames and also whether those features appear in games more or less often than in other pop-cultural products, thus shedding light on the specific potential role of videogame dialogue in unsupervised second language learning in leisure activities. Salience, on the other hand, refers to the acoustic or visual prominence of given linguistic elements, which makes them naturally stand out. In this respect, videogames – being multimedia products – offer an environment rich in acoustic and visual cues that may support noticing. Finally, when completing a task or achieving a goal that is meaningful to the learner depends on the correct comprehension or use of language, the learner’s focus on linguistic features increases. In videogames, this occurs, for example, when NPC dialogue serves a ludic function by providing instructions or suggestions to the player (Domsch, 2017; Dixon, 2022).

Given these premises, it comes to no surprise that the use of commercial off-the-shelf (COTS) videogames for language learning and their integration in formal learning contexts has attracted the attention of teachers and researchers (e.g. Kronenberg, 2012; Rankin & Shute, 2010). Learners themselves seem to be interested in using COTS videogames for language learning and empirically recognise the potential of gaming in language acquisition. An analysis of posts on two social communities, Quora and Reddit, showed that users frequently ask for suggestions on how to learn a foreign language using videogames and that the answers provided to this question demonstrate that users have rather clear ideas on the issue: “(1) users advocate a wide variety of vernacular game titles and genres for L2 learning, (2) as long they make the learner-player use language, (3) in a casual and enjoyable way. Users advise, (4) playing a game at the right proficiency level, (5) whose rules are not too unfamiliar, (6) and that include a lot of language use, (7) and features that provide time to read, re-read, listen, and re-listen. [...]” (Reinhardt, 2019, p. 34). Empirical research on the reasons why L2 English videogame players play off-the-shelf products in English confirms that a percentage of players engage in videogames with the specific aim to learn or improve their knowledge of the language. For example, in Italy, around 28% of university students

---

who report playing videogames in their spare time state that they do so to improve their L2 knowledge. This percentage increases to 41% when considering only students enrolled in foreign language curricula (Pavesi et al., forthcoming).

## **AIMS OF PRESENT STUDY AND RESEARCH QUESTIONS**

This study aims to contribute to establishing the role that scripted, non-player character dialogue in videogames can potentially play in the unsupervised acquisition of L2 English, compared to film dialogue. In particular, the study addresses the following research questions (RQs):

1. Do videogames and films expose learners to the same linguistic events, quantitatively and/or qualitatively? What, if any, grammatical features are a person more likely to encounter when playing videogames compared to when watching films?
2. How can such differences be explained?
3. How could the identification of differences between film and videogame dialogue inform language learning?

## **MATERIALS AND METHODS**

In this study a corpus of NPC dialogue in videogames was automatically analysed through a process called ‘POS tagging’ which matches each word in the corpus to a grammatical category, or part of speech (POS), and creates a list showing the frequency of those grammatical categories. Subsequently, the corpus was contrasted to three different reference corpora – also automatically tagged for parts of speech – to identify the grammatical categories that appear with significantly greater frequency in the corpus under investigation compared to the other sets of data. This process, called ‘key tag extraction’, is performed by the software by applying statistical methods (‘keyness measures’) which identify the linguistic features that are unusually frequent (or infrequent) in a specific corpus compared to another reference corpus. Keyness measures shed light on linguistic characteristics that are unique to the corpus under investigation.

Triangulation of the results of these automated analyses led to identifying a set of POS categories that can be reasonably considered statistically more frequent in NPC dialogue compared to film dialogue. The categories thus identified were further investigated in context, by automatically extracting and manually reading the lines (technically called ‘concordance lines’) containing the words matched by the software to each POS tag. This allowed us to refine the analyses and understand the functions that these parts of speech play in videogame dialogue.

The materials and methods used in this study are further detailed in the next few paragraphs.

### **The Corpora**

This study takes advantage of the Videogame Pilot Corpus (ViPiCo; Bianchi, 2024), summarized in Table 1. This corpus, comprising a total of 167,658 running words, includes scripted NPC dialogue from nine COTS videogames in a time span ranging from 1995 to 2021. The games are equally subdivided

across three game types: Action; Action-Adventure; and Role-Playing. We are aware that the size of this corpus is limited, but the ViPiCo was the only game dialogue corpus available to the author at the time of this research.<sup>7</sup> To overcome the limits of its dimension, only the features appearing in all the nine videogames were considered in the study.

*Table 1. ViPiCo – Details of the games comprising the corpus*

<b>Title</b>	<b>Year</b>	<b>Developer</b>	<b>Running words</b>
<i>Resident Evil Village</i>	2021	Capcom	10,324
<i>The Last of Us</i>	2013	Naughty Dog	27,301
<i>Devil May Cry V</i>	2019	Capcom	10,385
<i>Dead Space 2</i>	2011	Visceral Games	10,167
<i>Mafia Definitive Edition</i>	2020	Hangar 13	32,928
<i>Halo Infinite</i>	2021	343 Industries	14,535
<i>Bloodborne</i>	2015	From Software	12,429
<i>Chrono Trigger</i>	1995	Square	15,753
<i>Pokémon Black</i>	2010	Game Freak	33,836
ViPiCo			167,658

The ViPiCo was contrasted to the Anglophone section of the Pavia Corpus of Film Dialogue (Freddi & Pavesi, 2009),<sup>8</sup> which we shall here call UK-US FC. It currently comprises 34 American and British films covering a time range from 1996 to 2017 and totaling over 383,000 words. The UK-US FC has two relevant shortcomings: a miscellaneous composition, its films ranging from comedies to drama, from love to crime stories; and a relatively small size. For this reason, two other reference corpora were also used.

Specifically, since all the games in the ViPiCo have a strong action-oriented focus, we compared them to the seven films in the UK-US FC with action plots. These films are referred to hereafter as the Action Film Sub-Corpus (AFSC), as summarized in Table 2. This comparison allows us to exclude from the analyses any parts of speech that may appear key to the ViPiCo compared to the UK-US FC but are instead attributable to the differing focuses of the two corpora.

*Table 2. AFSC: Details*

<b>Title</b>	<b>Year</b>	<b>Director</b>	<b>Running words</b>
<i>Match Point</i>	2005	W. Allen	14,524
<i>Dead Man Walking</i>	1996	T. Robbins	13,149
<i>Michael Clayton</i>	2007	T. Gilroy	13,063
<i>Ocean's Eleven</i>	2001	S. Soderberg	10,390
<i>Thelma and Louise</i>	1991	R. Scott	11,191
<i>The Crying Game</i>	1992	N. Jordan	2,394
<i>Crash</i>	1994	P. Haggis & R. Moresco	10,326
Total			75,937

Finally, since the UK-US FC is too small to be considered representative of film dialogue in general, but film dialogue has been shown to compare to spontaneous interaction, we contrasted our videogame corpus with the BNC Sampler Spoken (BNC-SS), a representative subset of the spoken part of the British National Corpus (BNC).<sup>9</sup> The BNC is a collection of 100 million words, featuring samples of written and spoken language from diverse sources. It aims to provide a broad representation of British English from the late 20th century, encompassing both spoken and written forms. The BNC-SS comprises one million words of spoken texts recorded in educational, business, institutional and leisure contexts. In particular it includes the following types of texts: Lectures, talks, and educational demonstrations; News commentaries; Classroom interaction; Company talks and interviews; Trade union talks; Sales demonstrations; Business meetings; Consultations; Political speeches; Sermons; Public/government talks; Council meetings; Religious meetings; Parliamentary proceedings; Legal proceedings; Leisure speeches; Sports commentaries; Talks to clubs; Broadcast chat shows and phone-ins; and Club meetings.<sup>10</sup> Importantly, this part of the BNC does not include film or TV dialogue. Contrasting the ViPiCo with the BNC-SS allows us to exclude from the analyses any parts of speech which may appear key of the ViPiCo compared to the UK-US FC because of the size of the latter.

## Analytical Methods

The corpora were analysed and contrasted using Wmatrix 5 (Rayson, 2008), the most recent version of this online corpus analysis tool.<sup>11</sup> The primary affordances of Wmatrix include POS tagging based on the CLAWS7 tagset,<sup>12</sup> the presence of large reference corpora tagged with CLAWS, and the possibility to compare and contrast corpora at tag level. Another feature distinguishing Wmatrix from most corpus analysis tools is that it recognises multi-word-units and counts them as single tokens. Furthermore, a recently added functionality allows researchers to compare several files/corpora in a single click and establish the dispersion of individual items (e.g. word; POS tag). By cross-comparing the videogames in the corpus, it is possible to establish the distribution of POS tags.

The ViPiCo, the PCFD and the UK-US FC were uploaded in Wmatrix and POS tagged, while the BNC-SS was already available within Wmatrix as a reference corpus. The ViPiCo includes 157,089 tokens (words/multi-word units), 11,567 types and 167 POS tags. As Table 3 shows, 97 POS tags are common to all nine games, which corresponds to 58.1% of all POS tags.

*Table 3. Dispersion table of POS tags in the ViPiCo*

	POS TAGS	%
ViPiCo		
<i>Total</i>	<i>167</i>	<i>100</i>
Dispersion 9	97	58.1
Dispersion 8	13	7.8
Dispersion 7	5	3.0
Dispersion 6	9	5.4
Dispersion 5	7	4.2
Dispersion 4	6	3.6

continued on following page

Table 3. Continued

	POS TAGS	%
Dispersion 3	4	2.4
Dispersion 2	7	4.2
Dispersion 1	20	12.0

Next, the positive key POS tags characterizing the ViPiCo contrasted to each of the three reference corpora were extracted based on the log-likelihood (LL) value, a measure of statistical significance based on frequency; the threshold was set to 15.13, which corresponds to a very high level of significance ( $p < .001$ ). Log-likelihood is one of the most widely used methods in corpus linguistics to automatically compare corpora and has been shown to be suitable for identifying key words but also key grammatical categories (Rayson & Garside, 2000). Although effect size measures have recently been argued to be more valuable than significance values for corpus comparison (e.g. Gabrielatos & Marchi, 2012), for this particular study – aimed to identify the potential of videogames in language learning – frequency profiling was considered the best option, as frequency of exposure to a given linguistic event plays a crucial role in informal learning.

The resulting three lists were compared and the POS tags common to all of them were identified. These POS categories can be reasonably considered statistically more frequent in the ViPiCo compared to film dialogue generally. Furthermore, in order to overcome the limited size of our target corpus, we considered only those key items which are present in all the nine videogames (Dispersion 9). This is functional to the desire to try and identify those features which a player is bound to encounter regardless of the specific videogame they choose to play.

Finally, the concordance lines of the POS tags that are key in relation to all the three reference corpora and are present in all the videogames were manually analysed, to better understand the functions that these parts of speech play in videogame dialogue and their potential relevance for language learning. Thanks to the concordance feature, Wmatrix retrieves the lines of text containing the words that the software matched to each POS tag (known as ‘node words’) and lists them in order of appearance in the corpus, as shown in Figure 1.

Figure 1. Example of concordance lines of modal verbs (POS tag VM) in the ViPiCo as shown in Wmatrix.

```

along ... he saved me without sparing a thought for what might
strength, something incredible happened. Hoping that we would
: Oblivion came to an end. To regain his memories, Sora would
the Keyblade to create another Kingdom Hearts, and they would
willed through and continued onward. To ensure that Sora would
: strange messages kept appearing. Mickey and the others should
he bug in the data that had corrupted the journal. That should
ng in his heart. If he kept holding onto that hurt, he would
had left the mysterious journal entry, " Their hurting will
leathed Riku his Keyblade. Meanwhile, Ventus learned he would
r of darkness as he faces his own master. The Chi-blade would
ined attached to his armor. Even knowing that his heart would
that Sora and Riku still had a lot to learn before they could
y, Heartless appeared, and Aqua faced them knowing she 'd
: New Seven Hearts ... and in Toy Box, he learned hearts can
, and he was able to save Aqua from Vanitas. " Darkness will
hing 's wrong. Shut it down! This place ... This world will
ily possess the power of a princess of heart, your heart will
: resonate with a Keyblade wielder 's, and I believe you can
ive in. And yet, there are countless other worlds. You can
ntly in the night sky. Through that sea of stars, fate will
happen to him. Now it was my time to save him. As I wished for that 1 |
never be separated again, I gave Sora my lucky charm. The light eman 2 |
have to forget all about Namin. He fell into a deep sleep, but not b 3 |
harness it for their own gain. Realizing he was being used, Sora aim 4 |
make it to Organization XIII 's headquarters, Axel sacrificed himself 5 |
have been watching from the other side of the screen, but somewhere a 6 |
have been the end of this journey, but Mickey received word that a do 7 |
be dragged down into the darkness ... But Data-Sora realized that his 8 |
be mended when you return to end it. " Mickey and Data-Sora now knew 9 |
become the Chi-blade, and Master Eraqus tried to seal him away. Dete 10 |
form when pure light and pure darkness clashed. Ventus set out toward 11 |
be destroyed in the process, Ventus fought to stop the Chi-blade from 12 |
stop Master Xehanort, so he prompted them to take the Mark of Mastery 13 |
be trapped in the dark realm. Having fulfilled her duty as a Keyblade 14 |
reside where you least expect them. Meanwhile, Lea and I were hard a 15 |
prevail, and the light expire. " It all unfolded just as the Book of 16 |
be engulfed in darkness--everything you care about and everyone you lo 17 |
resonate with a Keyblade wielder 's, and I believe you can lead me to 18 |
lead me to that wielder 's world. That is the only possible way to sa 19 |
see them sparkle brilliantly in the night sky. Through that sea of st 20 |
guide your ship to where you are meant to be. However, if you arrive 21 |

```

The concordance lines were downloaded and saved in an Excel file for manual analysis. This analysis involved grouping the lines into homogeneous categories based on one or more of the following criteria: node word, e.g. a specific modal verb; co-text, e.g. the demonstrative immediately preceding a noun or an adjective; semantic traits of the node word, e.g. action verbs; tense, e.g. imperative vs. simple present; communicative function, e.g. deontic vs. epistemic. When necessary and feasible, the frequency of these homogenous categories was manually calculated as a percentage of the overall number of concordance lines in the POS tag or within the given subgroup.

## RESULTS AND DISCUSSION

Comparison of the ViPiCo to the UK-US FC identified 28 grammatical categories with statistically significantly higher frequency in the ViPiCo. Of these, as many as 25 are present in all the games; these tags are listed in alphabetical order in Table 4. In the table, column one lists Wmatrix's POS tags, columns two and three show each tag's raw and percentage frequencies in the ViPiCo, column four provides the LL values, and the last column offers the official explanation of the tag, as given in Wmatrix. The examples in parentheses are part of the tag's explanation and may not appear in the corpus under investigation.

*Table 4. ViPiCo vs. UK-US FC: Positive key POS tags appearing in all the nine games*

POS tag	Raw frequency	%	LL	Pos tag explanation
AT	5,488	3.49	233.22	article
CCB	902	0.57	19.98	adversative coordinating conjunction (but)
DAR	183	0.12	31.32	comparative after-determiner (e.g. more, less, fewer)
DB	718	0.46	33.99	before determiner or pre-determiner capable of pronominal function (all, half)
DD1	3,468	2.21	46.82	singular determiner (e.g. this, that, another)
DD2	349	0.22	15.80	plural determiner (these, those)
EX	426	0.27	20.34	existential there
IO	1,632	1.04	64.68	of (as preposition)
NN1	17,646	11.23	412.38	singular common noun
NN2	4,015	2.56	123.35	plural common noun
NNU	70	0.04	21.84	unit of measurement, neutral for number
PPHO2	479	0.30	50.25	3rd person plural objective personal pronoun (them)
PPIO2	481	0.31	127.02	1st person plural objective personal pronoun (US)
PPIS2	1,852	1.18	188.85	1st person plural subjective personal pronoun (we)
RL	1,869	1.19	133.88	locative adverb
RP	1,138	0.72	50.68	prep. adverb, particle (e.g. about, in)
RR22	40	0.03	161.02	general adverb
TO	2,688	1.71	20.35	infinitive marker
VBR	2,079	1.32	99.60	are

continued on following page

Table 4. Continued

POS tag	Raw frequency	%	LL	Pos tag explanation
VM	3,709	2.36	27.02	modal auxiliary
VV0	8,444	5.38	77.98	base form of lexical verb
VVI	6,958	4.43	100.15	infinitive (e.g. to give... It will work...)
VVN	2,124	1.35	21.18	past participle of lexical verb
VVZ	1,265	0.81	79.33	-s form of lexical verb (e.g. gives, works)
ZZ1	185	0.12	31.79	singular letter of the alphabet

Comparison between the ViPiCo and the AFSC identified 14 positive tags. Of these, 12 are present in all the games and are listed in alphabetical order in Table 5, alongside each tag's raw and percentage frequencies, the LL values and the official explanation of the tag, as given in Wmatrix.

Table 5. ViPiCo vs. AFSC: Positive key POS tags appearing in all the nine games

POS tag	Raw frequency	%	LL	POS tag explanation
AT	5,488	3.49	61.19	article (e.g. the, no)
DB	718	0.46	20.22	before determiner or pre-determiner capable of pronominal function (all, half)
IO	1,632	1.04	27.80	of (as preposition)
NN1	17,646	11.23	91.19	singular common noun (e.g. book, girl)
NN2	4,015	2.56	96.14	plural common noun (e.g. books, girls)
PPIO2	481	0.31	56.52	1st person plural objective personal pronoun (us)
PPIS2	1,852	1.18	22.99	1st person plural subjective personal pronoun (we)
RL	1,869	1.19	62.09	locative adverb (e.g. alongside, forward)
RP	1,138	0.72	38.30	prep. adverb, particle (e.g. about, in)
VM	3,709	2.36	78.89	modal auxiliary (can, will, would, etc.)
VV0	8,444	5.38	34.19	base form of lexical verb (e.g. give, work)
VVN	2,124	1.35	84.37	past participle of lexical verb (e.g. given, worked)

Finally, comparison between the ViPiCo and the BNC-SS identified 39 positive tags. Of these, 27 are present in all the games and are listed in alphabetical order in Table 6, alongside each tag's raw and percentage frequencies, the LL values and the official explanation of the tag, as given in Wmatrix.

Table 6. ViPiCo vs. BNC-SS: Positive key POS tags appearing in all the nine games

POS tag	Raw frequency	%	LL	Pos tag explanation
APPGE	2,853	1.82	589.77	possessive pronoun, pre-nominal (e.g. my, your, our)
IW	831	0.53	28.18	with, without (as prepositions)
JJ	7,561	4.81	362.52	general adjective

continued on following page

Table 6. Continued

POS tag	Raw frequency	%	LL	Pos tag explanation
JJT	136	0.09	31.86	general superlative adjective (e.g. oldest, best, strongest)
NN1	17,646	11.23	987.40	singular common noun (e.g. book, girl)
NP1	3,994	2.54	42.66	singular proper noun (e.g. London, Jane, Frederick)
PN1	1,217	0.77	701.79	indefinite pronoun, singular (e.g. anyone, everything, nobody, one)
PPHO1	707	0.45	16.59	3rd person sing. objective personal pronoun (him, her)
PPIO1	1,374	0.87	164.31	1st person sing. objective personal pronoun (me)
PPIO2	481	0.31	1070.57	1st person plural objective personal pronoun (us)
PPX1	148	0.09	323.75	singular reflexive personal pronoun (e.g. yourself, itself)
PPY	6,375	4.06	46.37	2nd person personal pronoun (you)
RL	1,869	1.19	806.86	locative adverb (e.g. alongside, forward)
RP	1,138	0.72	555.03	prep. adverb, particle (e.g. about, in)
RRQ	948	0.60	51.40	wh- general adverb (where, when, why, how)
RRR	268	0.17	90.69	comparative general adverb (e.g. better, longer)
TO	2,688	1.71	52.61	infinitive marker (to)
VB0	55	0.04	84.38	be, base form (finite i.e. imperative, subjunctive)
VBI	1,063	0.68	145.04	be, infinitive (To be or not... It will be ..)
VBM	910	0.58	51.56	am
VBR	2,079	1.32	20.96	are
VM	3,709	2.36	215.68	modal auxiliary (can, will, would, etc.)
VV0	8,444	5.38	157.59	base form of lexical verb (e.g. give, work)
VVD	2,702	1.72	28.25	past tense of lexical verb (e.g. gave, worked)
VVG	2,563	1.63	2781.69	-ing participle of lexical verb (e.g. giving, working)
VVI	6,953	4.43	66.24	infinitive (e.g. to give... It will work...)
VVZ	1,265	0.81	69.25	-s form of lexical verb (e.g. gives, works)

Tables 4, 5 and 6 have five tags in common. These shared key POS tags, summarized in Table 7, represent linguistic features that a player will encounter with much greater frequency in videogames than in film dialogue. Unlike the preceding tables, Table 7 does not include LL values, as these vary depending on the reference corpus.

Table 7. ViPiCo vs. UK-US FC, AFSC and BNC-SS: Positive key POS tags appearing in all the nine games

POS tag	Raw frequency	%	Pos tag explanation
NN1	17,646	11.23	singular common noun (e.g. book, girl)
RL	1,869	1.19	locative adverb (e.g. alongside, forward)
RP	1,138	0.72	prep. adverb, particle (e.g. about, in)
VM	3,709	2.36	modal auxiliary (can, will, would, etc.)
VV0	8,444	5.38	base form of lexical verb (e.g. give, work)

The following paragraphs analyse the concordance lines comprising the POS tags in Table 7.

## Singular Common Nouns

Manual analysis of the concordance lines comprising the tag NN1 shows that the overwhelming majority of singular common nouns are heads of specifying phrases introduced by the definite article (*the*), a demonstrative (e.g. *this*; *that*), a possessive adjective (e.g. *your*; *his*) or a Saxon genitive. The nouns themselves belong to a limited set of categories: character's roles (e.g. *the Pokemon*; *the gym leader*; *the Don*; *the boss*; *the champion*); characters 'congregations' (e.g. *the Elite Four*; *the Pokemon League*; *the Church of Unitology*); undefined but specific characters or groups of characters (e.g. *the guy who...*; *the rest of [us/you Spartans]*); places (*the gym*; *the Cathedral Ward*; *the Pokemon Center*; *the Ring*; *the city*); specific elements in the game (e.g. *the door*; *the way out*; *the power off/to [...]*; *the key to [...]*). In the comparatively very few cases when singular nouns belong to none of the above, they are part of phraseological units such as *what the fuck*, *who the fuck*, *the way of the world*. These noun phrases are characterized by simple forms of noun phrase modification, such as premodifying adjectives and nouns and postmodifying finite relative clauses and prepositional phrases, as found in spoken discourse (Biber et al., 2021).

From a learning standpoint, mastering noun phrases is a crucial skill for English learners across all proficiency levels. In particular, while the type of input provided by videogames is in line with the needs of the earliest stages of language learning (Biber et al., 2011), the correct use of definite noun phrases is particularly tricky for native speakers of languages that lack articles such as Russian, Korean, Japanese, or Chinese (e.g., Cho 2017; Huebner, 1983; Ionin et al., 2004, 2012; Snape, 2008; Thomas, 1989; Trenkic, 2008).

Furthermore, although the phrasal complexity of film dialogue generally aligns with that found in spontaneous conversation (Galiano, 2024), previous research on film dialogue has reported a higher frequency of pre- and post-modified noun phrases (Galiano, 2024) and greater use of demonstrative *this*, employed to foreground facts and objects relevant to the narrative or to illustrate the interlocutors' personal spheres and perspectives (Pavesi, 2020). These features imply greater specificity of film dialogue compared to unscripted conversation. Our findings, however, suggest that in-game scripted dialogue is characterized by a higher degree of specificity not only in comparison to real-world conversation (Dixon, 2022, p. 96) but also to film dialogue. For these reasons, input from videogame NPCs may be more beneficial than film input for acquiring and mastering the form and function of definite noun phrases.

Finally, our data confirm Dixon's (2022) observations that scripted dialogue in videogames is functional to providing clear and specific information about entities that are central to the quests and the main story of the game.

## Locative Adverbs

This category is dominated by adverbs *here* and *there*, totally comprising 69% of the concordance lines in this POS tag. Another relevant group of items is that of indefinite adverbs – which includes *anyplace*, *anywhere*, *everywhere*, *nowhere*, *someplace*, and *somewhere* – covering 6.2% of the concordance lines in this tag. Word *home*, used adverbially as in *I'm heading home*, amounts to 3.6% of the concordances, while *together* comprises 3.7% of the node words. Adverb *away* covers 2.5% of the concordances, while *ahead* and *inside* each represent 2.2% of the node words in this category. This means that as much

as 89.6% of this category is represented by only 13 words. Furthermore, 3.2% of the items in this list are phraseological units: *Away with you!* (5 items); *get away [with]* (22 items); *Here we go* (21 items); *there you are* (15 items); *there you go* (14 items); *here and there* (appearing only once); *there and then* (appearing only once). The remaining items are 35 different words, each appearing only a few times, composing 7.2% of the list.

Locative adverbs are considered distinguishing features of general conversation – compared to specialized types of conversation (Biber & Conrad, 2009, Ch. 4) – but have never been reported among the most prominent features of telecinematic dialogue. Consequently, NPC dialogue in videogames represents a privileged place for the observation of locative adverbs in context, albeit of a limited range of them. Their specific presence in videogame dialogue is most certainly due to the fact that in-game dialogue is required to support the ludic needs of the gameplay. In fact, while adverbs appear in the list of co-occurring features characterizing Interactive Discourse vs. Information on Specific Entities according to Dixon (2022), they do not appear as markers of Involved Production in Biber’s multidimensional scheme (Biber, 1988).

### Prepositional Adverbs/Particles

The tag RP includes the following words: *about, along, around, back, by, down, in, off, on, out, over, round, through, and up*. As expectable, these words are sometimes used to identify concrete physical and spatial locations or movements, as in examples [1]-[13]:

- [1] *Beware! There are many beasts about.*
- [2] *Trot along.*
- [3] *You ran around saying ‘Daddy’.*
- [4] *Stay back!*
- [5] *Pack a few bags. I’ll swing by after I go ‘round Paulie’s.*
- [6] *Stones of fire will rain down.*
- [7] *I think we got ta go in through the side.*
- [8] *Hands off, Mister.*
- [9] *Put these on.*
- [10] *And are you out looking for inspiration again?*
- [11] *That girl over there.*
- [12] *What about going through here?*
- [13] *Climb on up. Maybe we can cut through here.*

However, they also very frequently appear in figurative meaning, as in examples [14]-[26].

- [14] *What the hell was that all about?*
- [15] *You were at my side all along.*
- [16] *Stop fooling around.*
- [17] *Now, then, back to my introduction.*
- [18] *Have something to remember her by.*
- [19] *Are you two going to chase down Team Plasma like Cheren?*
- [20] *The verdict is in. Guilty!*

- [21] *I'm freezing my ass off.*  
 [22] *All systems on.*  
 [23] *Forgive me, I'm a bit out of sorts.*  
 [24] *It's not over yet.*  
 [25] *And you need to see this through.*  
 [26] *Live it up a little.*

This list includes not only instances of what Biber (1988) calls final prepositions, which are an indication of involved speech, but also locative adverbs, a marker of Interactive Discourse specific to NPC speech in videogames according to Dixon (2022). Once again, the scripted in-game dialogue between characters in videogames appears to be not only largely informal, but also different from both spontaneous conversation and film dialogue, most probably due to the player's need to interact with the ludic environment through the help and guidance of NPCs. Thus, NPC dialogue in videogames offers a unique opportunity to observe and potentially learn the meaning and use of prepositional adverbs/particles, which are notoriously difficult to acquire for L2 learners primarily due to their polysemous nature (Lam, 2009) and their sheer number (Koffi, 2010).

### Modal Auxiliaries

The POS category VM contains modal and semi-modal auxiliaries. The most frequent one is *will* (33.0%), followed by *can* (26.6%), *would* (15.2%), *could* (7%), *must* (5.7%), *should* (5.5%), *shall* (2.7%), *might* (2.3%), *may* (1.7%), *ought* (0.1%), *need* (0.1%) and *dare* (0.1%). Interestingly, this order of frequency does not fully coincide with the order of frequency that Leech (2003, Fig. 5, p. 231) found for modal verbs in contemporary spoken English in the ICEGB-mini-sp (International Corpus of English (Great Britain) "Mini-Corpus"; period of texts: 1990-92), where modal auxiliaries appeared in the following order: *will*; *can*; *would*; *should*; *could*; *might*; *may*; *must*; *shall*; *ought* (*need* appeared zero times). In particular, the positions of *shall* and *must* are markedly different, as they precede *may* and *might* in the ViPiCo, but follow them in the ICEGB-mini-sp.<sup>13</sup> Such a difference could theoretically be due to the different time spans of the two corpora (1995-2021 for the ViPiCo and 1990-92 for ICEGB-mini-sp), the nature of scripted/performed language, and/or the nature of videogames. However, the fact that a decline in the use of *must* and *shall* in both speech (-60.7% and -34.6%, respectively) and writing (-34.4% and -43.8%, respectively) can be observed from the 1960s to the 1990s (Leech, 2003, Fig. 5, p. 231 and Fig. 3, p. 228) suggests ruling out time span and scripted language as likely causes. Such marked trends in the evolution of language are not quickly reversed. On the other hand, it is much more likely that the percentages of *must* and *shall* be due to the nature of videogames. In fact, as the concordances show, *shall* appears in sentences with an archaic flavour, uttered by characters moving in magical and/or medieval-like worlds, as concordance lines [27]-[30] exemplify:

- [27] *We shan't allow thee to meddle with Lavos, Queen!*  
 [28] *Such impudence. Defiled, are We, yet still Queen. We shall not give audience to an ill-mannered beast.*  
 [29] *Frog! It was a fault of mine, which endangered the Queen. I shall depart for good. Crono, you hath potential to be a good swordsman!*

[30] *Or ... shall we employ the guillotine? You, the jury, shall decide his fate. Now, let us begin. Crono, you are hereby ordered to tell the truth!*

*Must*, too, appears in archaic-sounding lines, as in examples [31]-[33], but these lines do not cover a large percentage of instances of this modal verb. The higher presence of *must* in NPC talk compared to real life conversation is due to the directive force needed in in-game dialogue to guide the player in their quest, as in examples [34] and [35], or to explicate a character's standing with respect to the player or the general situation, as in examples [36] and [37]. In fact, while Leech (2003) observed 49% of deontic instances vs. 51% of epistemic instances of *must* in the ICEGB-mini-sp, the ViPiCo contains as many as 61.2% of deontic instances vs. only 38% of epistemic ones.

[31] *Ahh, well met. This is a most pleasing encounter. You see, I must depart erelong. My god is lost to me, so I must go away to find another.*

[32] *I tell you, I will not forget our adage. Fear the old blood. I must take my leave. Visitor ... I claim no subjects, but here lieth Our throne.*

[33] - *Dear Cyrus ... Thou must... think ill of me.*

- *On the contrary! You have come far, my friend.*

[34] *We must plan our escape! The Palace crumbles!*

[35] *We must liberate the Pokémon!*

[36] *You are a true hunter. Which is why I must stop you.*

[37] *Your blood ... is a very precious sacrifice. Eating you would do me no good. I must have every last drop ... to rule the Underworld ... as king !*

Furthermore, about 15.2% of the concordances in this POS tag revolve around the node word *would*, either identifying hypothetical situations, as shown in examples [38] and [39], or being used to express politeness, as in examples [40] and [41]:

[38] *I don't think I would make the cut*

[39] *Who would you bribe to get in here?*

[40] *I would appreciate if you had a Pokémon battle with me.*

[41] *If you want I would be glad if you could spend some more time with me.*

Considering all the lines comprising the VM category, the percentage of modals with an epistemic value is almost identical to that of modals with a dynamic value (29.3% and 29.4% respectively). Some examples of epistemic cases are:

[42] *There isn't no more survivors. It can't be.*

[43] *He will never hurt anyone. He is really very gentle.*

[44] *They must have called out every cop in the city. Come on, Tommy! Lose them.*

[45] *Try what you may. I shall not be stopped.*

[46] *Our prey might differ, but we are hunters, the both of us.*

Examples of dynamic cases include the following:

- [47] *I can't give you much in return, but I'll fight you with this Pokémon I raised real good!*  
 [48] *Griffin's Threat Sensor can help you locate Chak 'Lok.*  
 [49] *This is all we could get.*

Deontic instances amount to 25.2% of all modals, while 0.9% of cases are ambiguous. Instances of deontic modality include [50]-[54], while ambiguous cases are exemplified in [55] and [56]:

- [50] *Daddy, we can't leave him.* [indirect request]  
 [51] *Welcome to the world of Pokemon! You can call me Juniper.* [permission]  
 [52] *Don't mess with Crono, or you'll really be sorry.* [threat]  
 [53] *What is its extent? You must show me.* [order]  
 [54] *Hey what's in that hole? Should be take a look?* [suggestion]  
 [55] *Ooh, I'm making a mess of things ... I can't do this all that well!* [deduction or ability?]  
 [56] *Go back to patrols. We will get him next time.* [future tense or promise?]

Modal verbs – along with present tense, contractions, spoken verbs, first and second person pronouns, adverbs, the verb *be*, and the pronoun *it* – contribute to the positive end of Dixon's Interactive Discourse vs. Information on Specific Entities dimension. The current analysis, thus, confirms Dixon's (2002) observations that NPC lines inform the player of the volitions and abilities of the various characters, alert the player to possible danger, and support the player in choosing the best course of action through advice, promises, or simply by describing possible future consequences.

### Base Form of Lexical Verb

The VV0 tag suggests that players are largely exposed to the imperative form (57.3% of all lines in this tag). Imperatives in the ViPiCo are used to give instructions, orders, suggestions, and spurs (e.g., *abandon search and return to patrols*). They are less frequently used for requests (2.4% of all imperatives; e.g., *Please answer me*) and very rarely to express warnings (0.3% of all imperatives; e.g., *Beware the lady of the night; Fear the bell's toll*). A minimal number of imperative sentences in this category have *you* or *someone* as subject (e.g. *You dry your tears; Hey, someone give me a hand!*). On the other hand, there is comparatively ample presence of a few specific lexical forms: *Come on!* (3.4%); *Let...* (3.2%); *Look [!/,]* (2%). Such a large presence of imperatives is in keeping with the ludic function of NPC dialogue, i.e., giving instructions to the player (Domsch, 2017), and adds to the total amount of verbs performing a deontic function (Portner, 2007).

As many as 38.2% of the VV0 lines are examples of simple present tense, following subjects *I*, *we*, *you*, *they*, or specific plural names/nouns. A few verbs indicating mental activities frequently appear in the concordances: *know* (9.9% of all instances of present tense), *guess* (2.7%), *hope* (1.5%), and *wish* (1.3%). The subject is not always present, which is not uncommon in spoken English; this appears to be particularly true for the verb *guess*: e.g. *Guess this is why we're taking shitty protection jobs; Guess something serious is going down*. In both Dixon's (2022) and Biber's (1988) multidimensional analyses, the present tense is one of the co-occurring features – along with first and second person pronouns, and infinitives among others – that identify highly interactive and involved types of discourse.

The remaining items represent interesting, albeit minority, grammatical structures typical of informal discourse. About 2.1% of the occurrences in this tag are bare infinitives used in elliptical questions: e.g. *Care to respond, Chancellor?*; *Come to see me?*; *You hear me?*. About 0.1% are hidden conditionals of the type *You take one out, I'll take the other*; *You try it and you're dead*; *Play with fire and you get burned*. Finally, 2.3% of the node words are set phrases: *Welcome*; *Thank you*; *Excuse me*; *See you*; *Goddamit*; *Wonder what/if*.

At the syntactic level, two other features can easily be observed, cutting across the grammatical types listed above. One is the presence of colloquial forms where the verb phrase comprises two action verbs but the conjunction *and* is omitted: e.g. *I'll go check*; *wanna come check it?*; *Go ask him*. The other one is the sporadic presence of ungrammatical sentences, uttered by characters that the authors of the game wanted to mark as 'foreigners' (e.g. *We ask old king save shell*; *Old man breath, but dead on inside*; see Iaia, 2016).

At the lexical level, considering the entire VV0 category, a remarkable frequency of a few action verbs can be observed, declined in a wide range of phrasal constructions: the verb *take* (11% overall), sometimes followed by *away, back, down, off, out, in, along, up*, or used in the phrasal constructions *take a look, take care of, take it easy, take turns*; the verb *get* (5.2% overall), often followed by *back, down, out of, into, in, together, across, off, over, up*, or appearing in the phrasal construction *get the hell out of*; the verb *go* (3.6% overall), also followed by *after, ahead, around, away, back, by, down, for, in, on, through, under, up, with*; the verb *come* (3% overall), followed by *along, back, by, from, in, off, out, through, to, up*, but also appearing in the phrasal constructions *come to terms with*, and *come up with*; the verb *keep* (1.5% overall), frequently followed by a gerund, but also by particles *on, up*, or used in idiomatic expressions such as *keep quiet, keep an eye out, keep [one's] head down*; the verb *give* (1.3% overall), including instances followed by prepositions *back* and *up*, but also used in idiomatic expressions such as *give [someone] a break, give it a rest, give it a wide berth, give a hand, give and take*; the verb *look* (2.4%), also followed by *after, around, at, down, for, forward, into, like, out, to, up*; the verb *move* (0.6% overall), followed by *in, on, out, up*. The verb *need* is also very frequent in the VV0 group, covering 7.4% of all occurrences. Finally, the verb *see* is worth mentioning, as it covers 1.8% of concordance lines; a large percentage of its occurrences (23% of occurrences of the verb *see*) are due to the set phrase *See you [again/later/around]*. The remaining node words in this category are lexical verbs with very few occurrences each – excluding the verbs *be, have* and *do*, which are treated as separate tags by Wmatrix, regardless of auxiliary or lexical function.

## CONCLUSIONS AND FUTURE PERSPECTIVES

Our quantitative analyses showed that, although both film dialogue and videogame scripted dialogue tend to imitate spontaneous conversation, these two types of pop-cultural products do not equally expose learners to the same linguistic events (RQ1). In particular, five POS categories were found to be statistically more frequent in action videogames compared to film dialogue generally, as well as to real-life spontaneous conversation. Manual investigation of the concordance lines in the automatically identified POS categories refined the analyses, highlighting the lexical and grammatical features that videogame players are more likely to encounter, and learn, from videogames than from films. These can be summarised as follows: modal auxiliaries; base form of lexical verbs, primarily corresponding to imperatives and simple present tense (first and second person, singular and plural); a few highly-frequent action

verbs (*take; get; go; come; keep; give; look; move; need; and see*), declined in a wide range of phrasal constructions; a specific set of locative adverbs (*here, there, home, together, away, ahead and inside*, plus adverbs composed by prefix *any-* or suffix *-where*); prepositional particles *about, along, around, back, by, down, in, off, on, out, over, round, through, and up*; noun phrases introduced by definite determiners; elliptical questions; colloquial set phrases (such as *Away with you!; get away [with]; Here we go; there you are; there you go; there and then; what the fuck; who the fuck; the way of the world; give [someone] a break; give it a rest; give it a wide berth; give a hand; give and take; See you [again/later/around]*).

As argued in the Results and Discussion section, the lexico-grammatical features characterizing videogame dialogue compared to film dialogue are largely connected to the ludic function of videogames (Domsch, 2017), which is absent in films (RQ2). Players are required to interact with the ludic environment, and in doing so they are helped and guided by NPCs; NPCs give instructions to the player, alert them to possible danger, and support them in choosing the best course of action through advice, promises, or simply by describing possible future consequences (Dixon, 2022). These communicative functions are linguistically supported by the ample use of locative adverbs, prepositional adverbs/particles, modal verbs, and imperative forms. Furthermore, videogame dialogue shows a higher degree of specificity than film dialogue, which is linguistically achieved with the use of definite noun phrases providing clear and specific information about entities that are central to the quest and the main story of the game (see also Dixon 2022).

The identification of differences between film and videogame dialogue represents an important step not only towards a linguistic description of the two different types of pop-cultural artefacts, but also in a language learning perspective (RQ3): knowing what types of words and structures are available in a particular type of input and their respective functions provides learners and teachers with fundamental information for selecting the most appropriate input for a given learning objective or a specific language level and for predicting potential outcomes. When considered from an informal learning perspective, scripted in-game dialogue can be described as a ready-to-use source of conversational language exposing L2 learner-players to input that is largely comparable to spontaneous dialogue (Dixon 2022), in a highly engaging and motivating ludic environment (Reinhart & Thorne 2020) where situated learning is possible (Gee, 2006, 2008) and which generally favours language acquisition (De Wilde et al. 2019; De Wilde & Eyckmans 2017; Kuppens 2010; Sylvén & Sundqvist 2012; Sundqvist & Wikström 2015; Rød & Calafato 2024, in Calafato & Clausen 2024; Poole, 2020). Within this general backdrop, this study shows that compared to films – another highly appreciated form of input (Sundqvist, 2009; Verspoor et al., 2011) – videogames are potentially more useful for acquiring the set of lexical and grammatical features summarised above, which makes videogames more suitable for intermediate-to-advanced learner-users of English. This specific learning potential depends on two factors. One is the greater frequency of these features in videogames, frequency being a prerequisite for noticing and learning (Ellis, 2002; Gass & Mackey, 2002; Bybee, 2008). The other one is that understanding these linguistic elements is largely necessary for the players to orient themselves in the game and make choices in the gameplay – the ludic function (Domsch, 2017) of game dialogue – and when used to accomplish goals that are meaningful to the learner language becomes a powerful catalyst for learning (Loschky & Bley-Vroman 1993; Newton, 1995).

As a final adjunct, the current study shows the potential of corpus linguistics in comparing linguistic input from different sources and offers insight into the language of videogames. However, it has several limitations. Most notably, the ViPiCo is small, and includes only a subset of the many different existing types of videogames. Totally different videogame genres, such as sports games or graphic adventures,

---

could be expected to produce different results when compared to film dialogue. The film corpora used as terms of comparison are also small, which obliged us to triangulate results with data from a large corpus of natural spoken conversations. Comparison between much larger corpora of scripted dialogue in videogames, comprising sub-corpora each representative of a specific genre, and much larger corpora of film dialogue is needed to confirm the current results and provide insight into the specificities of dialogue in different game types. Furthermore, comparison between NPC talk and film dialogue at the level of semantic categories alongside POS categories could offer a more detailed and comprehensive view of the lexico-grammatical differences between the two types of pop-cultural products.

## ACKNOWLEDGMENT

This work was supported by the PRIN project, *The informalisation of English language learning through the media: Language input, learning outcomes and sociolinguistic attitudes from an Italian perspective*, funded by the Italian Ministry for University and Research (MUR) – grant 2020NNJTW3\_001.

## REFERENCES

- Álvarez-Bolado Sánchez, C., & Álvarez de Mon, I. (2019). Video games: A Lexical Approach. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 13–38). Bloomsbury. DOI: 10.5040/9781501338489.0007
- Balteiro, I. (2019). Lexical and Morphological Devices in Gamer Language in Fora. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 39–57). Bloomsbury. DOI: 10.5040/9781501338489.0008
- Bednarek, M. (2012). The language of fictional television. *Continuum*.
- Bednarek, M. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge University Press. DOI: 10.1017/9781108559553
- Bianchi, F. (2024). Videogame dialogue as input for language learning. An analysis of lexical coverage. *Lingue e Linguaggi*, 66, 241–266. DOI: 10.1285/i22390359v66p241
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. DOI: 10.1017/CBO9780511621024
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press. DOI: 10.1017/CBO9780511814358
- Biber, D., Gray, B., & Poonpon, K. (2011). Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development. *TESOL Quarterly*, 45(1), 5–35. DOI: 10.5054/tq.2011.244483
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2021). *Grammar of Spoken and Written English*. John Benjamins.
- Bogost, I. (2007). *Persuasive Games. The Expressive Power of Video games*. The MIT Press. DOI: 10.7551/mitpress/5334.001.0001
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In Robinson, P., & Ellis, N. C. (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 216–236). Routledge.
- Calafato, R., & Clausen, T. (2024). Vocabulary learning strategies in extramural English gaming and their relationship with vocabulary knowledge. *Computer Assisted Language Learning*, ●●●, 1–19. DOI: 10.1080/09588221.2024.2328023

- Campos-Pardillos, M. A. (2019). End-user Agreements in Video games: Plain English at Work in an Ideal Setting. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 116–135). Bloomsbury. DOI: 10.5040/9781501338489.0011
- Caruana, S. (2003). *Mezzi di comunicazione e input linguistico. L'acquisizione dell'italiano L2 a Malta*. Franco Angeli.
- Caruana, S. (2006). Trilingualism in Malta. Maltese, English and Italiano Televisivo. *International Journal of Multilingualism*, 3(3), 159–172. DOI: 10.2167/ijm024.0
- Cho, J. (2017). The acquisition of different types of definite noun phrases in L2-English. *The International Journal of Bilingualism*, 21(3), 367–382. DOI: 10.1177/1367006916629577
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 23(1), 171–185. DOI: 10.1017/S1366728918001062
- De Wilde, V., & Eyckmans, J. (2017). Game on! Young learners' incidental language learning of English prior to instruction. *Studies in Second Language Learning and Teaching*, 7(4), 673–694. DOI: 10.14746/ssl.2017.7.4.6
- Dixon, D. H. (2022). *The language in digital games: Register variation in virtual and real-world contexts* [Doctoral dissertation, Northern Arizona University]. ProQuest Dissertations & Theses.
- Dixon, D. H. (2024). Introducing the Single Player Offline Game Corpus (SPOC): A corpus of seven registers from digital role-playing games. *Corpora*, 19(1), 107–122. DOI: 10.3366/cor.2024.0300
- Domsch, S. (2017). Dialogue in video games. In Mildorf, J., & Thomas, B. (Eds.), *Dialogue across Media* (pp. 251–270). John Benjamins. DOI: 10.1075/ds.28.13dom
- Ellis, N. C. (2002). Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. DOI: 10.1017/S0272263102002024
- Ensslin, A. (2012). *The Language of Gaming*. Palgrave Macmillan. DOI: 10.1007/978-0-230-35708-2
- Ensslin, A., & Finnegan, J. (2019). Bad Language and Bro-up Cooperation in Co-sit Gaming. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 139–156). Bloomsbury. DOI: 10.5040/9781501338489.0013
- Fernández-Vara, C. (2014). *Introduction to Game Analysis*. Routledge. DOI: 10.4324/9780203794777
- Fizek, S. (2012). *Pivoting the Player. A methodological toolkit for player character research in offline role-playing games*. ProQuest Dissertations Publishing. <https://www.proquest.com/openview/385e0788aac018f08ac6021469e42fc1/1?pq-origsite=gscholar&cbl=51922> (12.7.2023).
- Forchini, P. (2012). *Movie language revisited: Evidence from multi-dimensional analysis and corpora*. Peter Lang. DOI: 10.3726/978-3-0351-0325-0
- Formentelli, M. (2014). Vocatives Galore in Audiovisual Dialogue: Evidence from a corpus of American and British films. *English Text Construction*, 7(1), 53–83. DOI: 10.1075/etc.7.1.03for

- Formentelli, M., & Zago, R. (2024). Complexity matters in TV dialogue as language input. A corpus-based description of clausal and phrasal patterns. *Lingue e Linguaggi*, 66, 267–293. DOI: 10.1285/i22390359v66p267
- Freddi, M. (2011). A Phraseological approach to film dialogue: Film stylistics revisited. *Yearbook of Phraseology*, 2, 137–63. <https://doi.org/DOI: 10.1515/9783110236200.137>
- Freddi, M., & Pavesi, M. (2009). The Pavia Corpus of Film Dialogue: Research rationale and methodology. In Pavesi, M., & Freddi, M. (Eds.), *Analysing audiovisual dialogue. Linguistic and translational insights* (pp. 95–100).
- Gabrielatos, C., & Marchi, A. (2012, September 13–14). Keyness: Appropriate metrics and practical issues [Conference presentation]. CADS International Conference. Corpus-assisted Discourse Studies: More than the sum of Discourse Analysis and computing? Bologna, Italy.
- Galiano, L. (2024). Noun phrase complexity in film dialogue as input for second-language learning. A corpus-based study from a register-functional perspective. *Lingue e Linguaggi*, 66, 295–318. DOI: 10.1285/i22390359v66p295
- Gass, S. (1997). *Input, interaction, and the second language learner*. Erlbaum.
- Gass, S. M., & Mackey, A. (2002). Frequency Effects and Second Language Acquisition. A Complex Picture? *Studies in Second Language Acquisition*, 24(2), 249–260. DOI: 10.1017/S0272263102002097
- Gee, J. P. (2006). Are video games good for learning? *Nordic Journal of Digital Literacy*, 1(3), 172–183. DOI: 10.18261/ISSN1891-943X-2006-03-02
- Gee, J. P. (2008). Learning theory, video games and popular culture. In Drotner, K., & Livingstone, S. (Eds.), *The International Handbook of Children, Media and Culture* (pp. 196–213). SAGE Publications. DOI: 10.4135/9781848608436.n13
- Gee, J. P. (2014). *Unified Discourse Analysis: Language, Reality, Virtual Worlds and Video Games*. Routledge. DOI: 10.4324/9781315774459
- Ghia, E. (2014). ‘That is the question’: Direct interrogatives in English film dialogue and dubbed Italian. In Pavesi, M., Formentelli, M., & Ghia, E. (Eds.), *The languages of dubbing: Mainstream audiovisual translation in Italy* (pp. 57–88). Peter Lang.
- Gledhill, C. (2019). Phraseology and Lexico-grammatical Patterns in Two Emergent Paragame Genres: Video game Tutorials and Walkthroughs. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 58–86). Bloomsbury. DOI: 10.5040/9781501338489.0009
- Graham, S. L., & Dutt, S. (2019). “Watch the Potty Mouth”: Negotiating Impoliteness in Online Gaming. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 201–223). Bloomsbury. DOI: 10.5040/9781501338489.0017
- Hancock, M. (2019). Playing It By the Book: Instructing and Constructing the Player in the Video game Manual Paratext. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 288–303). Bloomsbury. DOI: 10.5040/9781501338489.0021

- Heidt, J., Pinchbeck, G. G., & Rodgers, M. P. H. (2023). The Good Gaming (GG) list: Key vocabulary in videogames. In Reynolds, B. L. (Ed.), *Vocabulary learning in the wild* (pp. 143–161). Springer. DOI: 10.1007/978-981-99-1490-6\_5
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Karoma.
- Iaia, P. (2016). *Analysing English as a lingua franca in video games. Linguistic features, experiential and functional dimensions of online and scripted interactions*. Peter Lang.
- Iaia, P. (2021). Towards a ‘COOPING’ Model for the Investigation of Gamers’ Online Conversations in English. *Iperstoria*, 17, 222–240. DOI: 10.13136/2281-4582/2021.i17.902
- Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2012). That’s not so different from the: Definite and demonstrative descriptions in second language acquisition. *Second Language Research*, 28(1), 69–101. DOI: 10.1177/0267658311432200
- Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1), 3–69. DOI: 10.1207/s15327817la1201\_2
- Kiourti, E. (2019). “Shut the Fuck up Re! Plant the Bomb Fast!”: Reconstructing Language and Identity in First-person Shooter Games. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 157–177). Bloomsbury. DOI: 10.5040/9781501338489.0014
- Koffi, E. (2010). *Applied English syntax: Foundations for word, phrase, and sentence analysis*. Kendall Hunt Publishing Company.
- Kronenberg, F. A. (2012). Selection Criteria for Commercial Off-the-Shelf (COTS) Video Games for Language Learning. *IALLT Journal of Language Learning Technologies*, 42(2), 52–78. DOI: 10.17161/iallt.v42i2.8512
- Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, 35(1), 65–85. DOI: 10.1080/17439880903561876
- Lam, Y. (2009). Applying cognitive linguistics to teaching the Spanish prepositions *por* and *para*. *Language Awareness*, 18(1), 2–18. DOI: 10.1080/09658410802147345
- Leech, G. (2003). Modality on the move: The English modal auxiliaries 1961–1992. In F. Roberta M. Krug, & F. Palmer (Ed.), *Modality in Contemporary English* (pp. 223–240). De Gruyter.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In Crookes, G., & Gass, S. (Eds.), *Tasks and Language Learning: Integrating Theory and Practice* (pp. 123–167). Multilingual Matters.
- Muñoz, C. (2020). Boys like games and girls like movies. Age and gender differences in out-of-school contact with English. *Revista Española de Lingüística Aplicada*, 33(1), 172–202. DOI: 10.1075/res-la.18042.mun
- Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, 11(2), 159–177. DOI: 10.1177/026765839501100207

- Paul, C. A. (2012). *Wordplay and the Discourse of Video Games. Analysing Words, Design and Play*. Routledge. DOI: 10.4324/9780203124031
- Pavesi, M. (2013). This and That in the Language of Film Dubbing: A Corpus-based Analysis. *Meta*, 58(1), 107–137. DOI: 10.7202/1023812ar
- Pavesi, M. (2016). Formulaicity in and across film dialogue: Clefts as translational routines. *Across Languages and Cultures*, 17(1), 99–121. DOI: 10.1556/084.2016.17.1.5
- Pavesi, M. (2020). ‘I shouldn’t have let this happen’. Demonstratives in film dialogue and film representation. In Hoffmann, C., & Kirner-Ludwig, M. (Eds.), *Telecinematic stylistics* (pp. 19–38). Bloomsbury. DOI: 10.5040/9781350042889.0007
- Pavesi, M. (2022). Corpora and the language of films: exploring dialogue in English and Italian. In O’Keeffe, A., & McCarthy, M. J. (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed., pp. 547–561). Routledge. DOI: 10.4324/9780367076399-38
- Pavesi, M., Bianchi, F., Bruti, S., & Zago, R. (forthcoming). In contact with English, Informal media-driven learning among Italian university students. *Carocci*.
- Peters, E. (2018). The effect of out-of-school exposure to English language media on learners’ vocabulary knowledge. *ITL International Journal of Applied Linguistics*, 169(1), 142–167. DOI: 10.1075/itl.00010.pet
- Peterson, M. (2010). Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning*, 23(5), 429–439. DOI: 10.1080/09588221.2010.520673
- Peterson, M. (2011). Digital gaming and second language development: Japanese learners interactions in a MMORPG. *Digital Culture & Education*, 3(1), 56–73. [https://www.digitalcultureandeducation.com/cms/wp-content/uploads/2011/04/dce1048\\_peterson\\_2011.pdf](https://www.digitalcultureandeducation.com/cms/wp-content/uploads/2011/04/dce1048_peterson_2011.pdf)
- Peterson, M. (2012). Learner interaction in massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL*, 24(03), 361–380. DOI: 10.1017/S0958344012000195
- Poole, F., & Clarke-Midura, J. (2020). A Systematic Review of Digital Games in Second Language Learning Studies. *International Journal of Game-Based Learning*, 10(3), 1–15. DOI: 10.4018/IJGBL.2020070101
- Portner, P. (2007). Imperatives and modals. *Natural Language Semantics*, 15(4), 351–383. DOI: 10.1007/s11050-007-9022-y
- Quaglio, P. (2008). Television dialogue and natural conversation: Linguistic similarities and functional differences. In Annelie, Ä., & Reppen, R. (Eds.), *Corpora and discourse: The challenges of different settings* (pp. 198–210). John Benjamins. DOI: 10.1075/scl.31.12qua
- Quaglio, P. (2009a). *Television dialogue: The sitcom Friends vs. natural conversation*. John Benjamins. DOI: 10.1075/scl.36
- Quaglio, P. (2009b). Vague language in the situation comedy Friends vs. natural conversation. In Fredi, M., & Pavesi, M. (Eds.), *Analysing audiovisual dialogue: Linguistic and translational insights* (pp. 75–91). Clueb.

- Rankin, Y. A., McNeal, M., & Shute, M. W. (2008). User centered game design: Evaluating massive multiplayer online role playing games for second language acquisition. In *Sandbox '08: Proceedings of the 2008 ACM SIGGRAPH symposium on Video games* (pp. 43–49). Association for Computing Machinery. DOI: 10.1145/1401843.1401851
- Rankin, Y. A., & Shute, M. W. (2010). Re-purposing a recreational video game as a serious game for second language acquisition. In Kaufman, D., & Sauvé, L. (Eds.), *Educational gameplay and simulation environments: Case studies and lessons learned* (pp. 178–194). Information Science Reference.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549. DOI: 10.1075/ijcl.13.4.06ray
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In Kilgarriff, A., & Berber Sardinha, T. (Eds.), *Proceedings of the workshop on Comparing Corpora. Held in conjunction with The 38th annual meeting of the Association for Computational Linguistics* (pp. 1–6). Association for Computational Linguistics., <https://aclanthology.org/volumes/W00-09/>
- Reinhardt, J. (2019). *Gameful second and foreign language teaching and learning: Theory, research, and practice*. Palgrave Macmillan. DOI: 10.1007/978-3-030-04729-0
- Reinhardt, J., & Thorne, S. L. (2020). Digital Games as Language- Learning Environments. In Plass, J. L., Mayer, R. E., & Homer, B. D. (Eds.), *Handbook of game-based learning* (pp. 409–435). The MIT Press.
- Rennick, S., & Roberts, S. (2024). The Video Game Dialogue Corpus. *Corpora*, 19(1), 93–106. DOI: 10.3366/cor.2024.0299
- Rød, A. J., & Calafato, R. (2024). Exploring extramural English: Impacts, integration and future directions. In Skulstad, A. S. (Ed.), *Current issues in English teaching* (pp. 117–140). Fagbokforlaget.
- Rodgers, M. P. H., & Heidt, J. (2021). Levelling up comprehensible input and vocabulary learning. The lexical profile of videogames. In Werner, V., & Tegge, F. (Eds.), *Pop culture in language education. Theory, research, practice* (pp. 215–227). Routledge.
- Rudge, L. A. (2019). “I Cut It and I ... Well Now What?”: (Un)Collaborative Language in Timed Puzzle Games. In Ensslin, A., & Balteiro, I. (Eds.), *Approaches to Video game Discourse. Lexis, interaction, textuality* (pp. 178–200). Bloomsbury. DOI: 10.5040/9781501338489.0015
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/DOI: 10.1093/applin/11.2.129>
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In Schmidt, R. (Ed.), *Attention and awareness in foreign language learning* (pp. 1–65). University of Hawaii Press.
- Schmidt, R. (2001). Attention. In Robinson, P. (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. DOI: 10.1017/CBO9781139524780.003

- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721–737). National University of Singapore, Center for Language Studies.
- Schmidt, R., & Frota, S. N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In Day, R. (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Newbury House.
- Snape, N. (2008). *The acquisition of the English determiner phrase by L2 learners: Japanese and Spanish*. VDM Verlag.
- Stamenković, D. (2023). The stylistic journey of a video game. A diachronic approach to multimodality in the Football Manager series. In Schubert, C., & Werner, V. (Eds.), *Stylistic Approaches to Pop Culture* (pp. 227–246). Routledge.
- Stamenković, D., & Jačević, M. (2019). Video Games and Multimodality: Exploring Interfaces and Analyzing Video Game Screens Using the GeM Model. In Wildfeuer, J., Pflaeging, J., Bateman, J., Seizov, O., & Tseng, C.-I. (Eds.), *Multimodality. Disciplinary Thoughts and the Challenge of Diversity* (pp. 277–294). De Gruyter. DOI: 10.1515/9783110608694-011
- Sundqvist, P. (2009). *Extramural English Matters. Out-of-School English and Its Impact on Swedish Ninth Graders' Oral Proficiency and Vocabulary* [Doctoral dissertation, Karlstad University Studies]. <https://www.diva-portal.org/smash/get/diva2:275141/FULLTEXT03>
- Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, 51, 65–76. DOI: 10.1016/j.system.2015.04.001
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302–321. DOI: 10.1017/S095834401200016X
- Thomas, M. (1989). The acquisition of English articles by first-and second-language learners. *Applied Psycholinguistics*, 10(3), 335–355. DOI: 10.1017/S0142716400008663
- Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition*, 11(1), 1–18. DOI: 10.1017/S1366728907003185
- Veirano Pinto, M. (2014). Dimensions of variation in North American movies. In Berber Sardinha, T., & Veirano Pinto, M. (Eds.), *Multi-dimensional analysis, 25 years on. A tribute to Douglas Biber* (pp. 109–147). John Benjamins. DOI: 10.1075/scl.60.04vei
- Verspoor, M. H., de Bot, K., & van Rein, E. (2011). English as a foreign language. The role of out-of-school language input. In A. De Houwer, & A. Wilton (Eds.), *English in Europe Today. Sociocultural and Educational Perspectives* (pp. 148-166), John Benjamins. DOI: 10.1075/aals.8.10ver
- Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(4), 407–427. DOI: 10.1093/applin/amp010
- Werner, V. (2021). Text-linguistic analysis of performed language: Revisiting and re-modeling Koch & Oesterreicher. *Linguistics*, 59(3), 541–575. DOI: 10.1515/ling-2021-0036

Werner, V. (2022). Pop Cultural Linguistics. In *Oxford research encyclopedia of linguistics* (n.p.). Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.999

Zago, R. (2016). *From originals to remakes. Colloquiality in English film dialogue over time*. Bonanno Editore.

Zago, R. (2021). Film discourse. In Friginal, E., & Hardy, J. A. (Eds.), *The Routledge handbook of discourse approaches to discourse analysis* (pp. 168–182). Routledge.

## ADDITIONAL READING

Stamenković, D., Jačević, M., & Wildfeuer, J. (2016). The persuasive aims of Metal Gear Solid: A discourse theoretical approach to the study of argumentation in videogames. *Discourse, Context & Media*, 15, 11–23. DOI: 10.1016/j.dcm.2016.12.002

Werner, V., & Tegge, F. (Eds.). (2021). *Pop culture in language education. Theory, research, practice*. Routledge.

## KEY TERMS AND DEFINITIONS

**Epistemic Modality:** Epistemic modality is said to apply when a word or phrase expresses how confident a speaker or writer is about the claims they make.

**Deontic Modality:** Deontic modality is said to apply when a word or phrase is used to perform linguistic acts such as giving an order or suggestion, threatening, promising, etc.

**Dynamic Modality:** Dynamic modality applies to modal words indicating ability or possibility (typically can, be able to, will).

**Multidimensional Analysis:** A statistical process that groups data into categories and analyses the co-occurrence of those categories; multidimensional analysis was first introduced in linguistics by Douglas Biber and used to distinguish different registers and text types in English.

**Tagging:** Automatic process by which the words in a corpus are matched to specific analytical categories, such as parts of speech or semantic fields.

**Key Item:** An item, e.g. word or part-of-speech category, that appears significantly more frequently (positive key item) or less frequently (negative key item) in a set of data compared to a different set of data.

**Videogame:** Electronic game featuring among the products of pop culture.

## ENDNOTES

<sup>1</sup> For example, Queen observes that a change in funding models impacts on the use of language variation (as cited in Bednarek, 2018, p. 6), while Epstein refers that dialogue between two people is easier and cheaper than multiparty conversations (as cited in Bednarek, 2018, p. 14).

- 
- <sup>2</sup> Biber (1988) was the first to introduce multidimensional analysis in corpus linguistics and used it to compare different written and spoken registers to each other and to identify the grammatical features of spontaneous conversation. He considered the following dimensions: Involved versus Informational Production; 2. Narrative versus Non-Narrative Concerns; 3. Explicit versus Situation-Dependent Reference; 4. Overt Expression of Persuasion; 5. Abstract versus Non-Abstract Information; 6. On-Line Informational Elaboration.
- <sup>3</sup> More recent productions are marked by even greater involvement than those of the 1950s and 1960s (Zago, 2016).
- <sup>4</sup> For a complete description of this corpus see Dixon (2022), but also Dixon (2024).
- <sup>5</sup> Dixon (2022) also analysed the written text of the videogames in his corpus, divided into registers.
- <sup>6</sup> Studies contradicting this finding also exist (e.g. Peters 2018; Muñoz 2020). Inconsistent findings in the literature may be explained by factors such as participants' age, language proficiency, or type of game played (Peters, 2018; Muñoz, 2020).
- <sup>7</sup> The analyses were carried out in 2023 and the results were presented at a conference in Lecce, Italy, on 8-9 February 2024. In 2024, a large-scale open access corpus of videogame dialogue appeared (Rennick & Roberts, 2024).
- <sup>8</sup> The PCFD was conceived as a monitor/open corpus. The official description of this corpus is available at <https://linguisticapavia.cdl.unipv.it/it/il-corso/progetti-diricerca>. I heartfully thank Maria Pavesi for giving me direct access to the individual files comprising the corpus and allowing me to run them in Wmatrix.
- <sup>9</sup> <http://www.natcorp.ox.ac.uk/>
- <sup>10</sup> <http://www.natcorp.ox.ac.uk/archive/worldURG/design.xml?ID=spodes>.
- <sup>11</sup> Version 6 of Wmatrix also exists but is still in beta testing.
- <sup>12</sup> <https://ucrel.lancs.ac.uk/claws/>
- <sup>13</sup> Statistical comparison between the data from the two corpora is not possible, since Leech (2003) does not provide percentage counts for these verbs.