Teachers by accident: on the potential of non-educational video games for incidental foreign language vocabulary acquisition

Tim Vincent Kassenberg

S1770942

MA in Applied Linguistics
Faculty of Liberal Arts
University of Groningen

Supervisors:
Dr. Marjolijn H. Verspoor
Dr. Steven L. Thorne

28-06-2013
Acknowledgements

There are several people who deserve my gratitude for helping me with my thesis. First of all, I thank my supervisors for their invaluable patience, support and insights during my work. I also thank my fellow applied linguistics student Sara Menegoni for her help with the Spanish words and phrases used in the tests. Additionally, my thanks are due to my good friends Jelle and Arjen, who have spent a great many hours with me in the university library, writing or discussing our individual theses and providing me with advice and encouragement when I needed it. Finally I thank my parents for their continuous support during this long and challenging period.
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>First language</td>
</tr>
<tr>
<td>L2</td>
<td>Second language</td>
</tr>
<tr>
<td>FL</td>
<td>Foreign language</td>
</tr>
<tr>
<td>TL</td>
<td>Target language</td>
</tr>
<tr>
<td>NS(s)</td>
<td>Native speaker(s)</td>
</tr>
<tr>
<td>NES(s)</td>
<td>Native English speaker(s)</td>
</tr>
<tr>
<td>LL(s)</td>
<td>Language learner(s)</td>
</tr>
<tr>
<td>NNS(s)</td>
<td>Non-native speaker(s)</td>
</tr>
<tr>
<td>ESLL(s)</td>
<td>English second language learner(s)</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a foreign language</td>
</tr>
<tr>
<td>ESL</td>
<td>English as a second language</td>
</tr>
<tr>
<td>DUB</td>
<td>Dynamic usage-based (approach)</td>
</tr>
<tr>
<td>DST</td>
<td>Dynamic systems theory</td>
</tr>
<tr>
<td>ILH</td>
<td>Involvement load hypothesis</td>
</tr>
<tr>
<td>Y/N test</td>
<td>Yes/No checklist test / checklist questionnaire</td>
</tr>
<tr>
<td>VG(s)</td>
<td>Video game(s)</td>
</tr>
<tr>
<td>COTS</td>
<td>Commercial off-the-shelf</td>
</tr>
<tr>
<td>MMO(RPG)</td>
<td>Massively multi-player online (role playing game)</td>
</tr>
<tr>
<td>FPS</td>
<td>First person shooter</td>
</tr>
<tr>
<td>HUD</td>
<td>Heads-up display</td>
</tr>
<tr>
<td>NPC(s)</td>
<td>Non-playing character(s)</td>
</tr>
<tr>
<td>VHF</td>
<td>Very high frequency</td>
</tr>
<tr>
<td>HF</td>
<td>High frequency</td>
</tr>
<tr>
<td>TF</td>
<td>Threshold frequency</td>
</tr>
<tr>
<td>LF</td>
<td>Low frequency</td>
</tr>
<tr>
<td>PW</td>
<td>Pseudo-word</td>
</tr>
<tr>
<td>Isdt</td>
<td>Index of signal detection theory</td>
</tr>
</tbody>
</table>
Table of contents

Abstract 6
Introduction 7
Background 8
Chapter 1: Approaches to language acquisition 8
  Chapter 1.1: A dynamic usage-based approach to language acquisition 8
  Chapter 1.2: A dynamic systems theory approach to language acquisition 9
Chapter 2: Factors in foreign language vocabulary acquisition 10
  Chapter 2.1: The nature of the lexicon 10
  Chapter 2.2: The role of frequency in vocabulary acquisition 12
  Chapter 2.3: The role of context in vocabulary acquisition 14
  Chapter 2.4: Resources for vocabulary acquisition 16
  Chapter 2.5: Dynamic language development 17
Chapter 3: Video games as foreign language vocabulary acquisition environments 19
  Chapter 3.1: Terminology and typology: educational versus non-educational video games 19
  Chapter 3.2: Factors facilitating incidental vocabulary acquisition from COTS video games 20
  Chapter 3.3: Factors inhibiting incidental vocabulary acquisition from COTS video games 23
  Chapter 3.4: Statement of purpose 26
Chapter 4: Methodology 28
  Chapter 4.1: Participants 28
  Chapter 4.2: Materials 29
  Chapter 4.3: Procedures 35
  Chapter 4.4: Design and analysis 35
Chapter 5: Results 38
  Chapter 5.1: Quantitative data: I_{cat} and chunk completion scores 38
  Chapter 5.2: Qualitative data: retrospective questionnaires 41
Chapter 6: Discussion 43
  Chapter 6.1: Incidental vocabulary acquisition from COTS video games 43
  Chapter 6.2: Factors facilitating incidental FL vocabulary acquisition from COTS video games 43
  Chapter 6.3: Factors inhibiting incidental FL vocabulary acquisition from COTS video games 44
  Chapter 6.4: Limitations of the study 45
Chapter 7: Conclusion 46
References 47
Abstract

This explorative study discusses several factors influencing incidental vocabulary acquisition in general and the potential of video games for foreign language vocabulary learning in particular. Based on previous research into this topic, seven participants were tested on their receptive vocabulary size in foreign language Spanish, in terms of single lexical items and chunks. The study used a pre-test/post-test/delayed post-test design before and after playing a Spanish language video game, supplemented by a retrospective questionnaire. The results showed only limited statistical support for the hypotheses stating that receptive vocabulary knowledge increases between tests. Additionally, the qualitative retrospective data showed several common themes, most notable of which are problematic cognitive load and consequent incomprehensibility of the target language, and the benefits of situated meaning to incidental vocabulary acquisition from video games.
Introduction

Since the creation of Pong in 1972, video games have followed the path of computers in rapidly becoming more complex, facilitating constantly better visual realism, and becoming more widespread. This is reflected in surveys conducted around the world, showing that many millions of people in Europe, China and the United States spending around 20 hours per week (or far more!) on playing video games (McGonigal 2011).

All of this means that today, millions of people around the world are frequently exposed to an interactive medium which continues to become more complex and realistic, both in terms of graphical realism and its story-telling qualities. Therefore, the author of this study, along with many other researchers whose work is discussed here, is of the opinion that the potential of video games to provide significant opportunities for learning to gamers, merits immediate investigation.

In this study, the question of whether or not foreign language vocabulary can be incidentally acquired from video games is investigated. First, several of the many factors found to affect vocabulary acquisition in previous research are discussed. Next, these factors are discussed in relation to several unique characteristics and affordances of video games for language learning research has been able to identify. Finally, several hypotheses concerning the incidental acquisition of vocabulary from video games are formulated and tested, using a variety of methods in order to answer the general research question immanent in this study: is incidental acquisition of foreign language vocabulary from video games possible?
Chapter 1: Approaches to language acquisition

Introduction

The study of second and foreign language acquisition has a long history and has been viewed from different perspectives throughout the history of the field. The most important distinction is that between the traditions of cognitive and generative linguistics. The many approaches usually grouped under the name ‘generative linguistics’ focus on finding a minimalist system underlying language. In contrast, the approaches within cognitive linguistics have a more maximalist, non-reductive view of language, seeing it as an ‘inventory of forms’ rather than a system of parameters and rules (Langacker 2000). The purpose of this thesis is not to explain the many differences between these traditions. However, it is important to make clear that the process of language acquisition will be viewed from the perspective of cognitive linguistics, and two complementary approaches seen as part of this tradition: usage-based linguistics and dynamic systems theory.

1.1: A Dynamic Usage Based approach to language acquisition

The first of these approaches is the Dynamic Usage Based (DUB) approach to language acquisition proposed by Langacker (2000, 2009) and others. This approach is based on the same assumptions as other approaches in the cognitive linguistic tradition, the most important of which is the basic concept that language acquisition is achieved through the use of several abilities and processes inherent in human cognition. Consequently, in contrast to the fundamental arguments underlying generative linguistics, no specialized and separate language learning faculty in human cognition is required (Langacker 2000). Instead, more general cognitive capacities such as perception, categorization, and memory are viewed as central to language acquisition. Consequently, what is acquired can be viewed as a large system of conventional units in each language domain (i.e. semantics, phonology, lexicon, morphology and syntax) (Langacker 2000). Thus, the language system in the DUB approach is seen as a collection of forms, units and schemas, instead of a system of rules and parameters.

In order to build this collection there are several cognitive processes which allow learners to acquire language from usage events. The first is ‘entrenchment’, denoting the process by which language forms (words, phrases, sentences etc.) encountered in usage events leave traces which facilitate their recurrence (Langacker 2009). Combinations of words so thoroughly entrenched that they no longer require conscious attention to use are given the status of ‘unit’. The second process is ‘abstraction’, in which the common aspects of usage events are reinforced and non-recurring aspects
are thus ‘filtered out’. This leads to schematization, with a ‘schema’ defined as “[...] the commonality that emerges from distinct structures when one abstracts away from their points of difference by portraying them with lesser precision and specificity.” (Langacker 2000:4). Consequently, these schemas operate at many different levels in the language system and together form the collection of conventional units that make up that system. In turn, schemas make it possible for learners to produce novel units and thus be linguistically creative. Once the different elements of usage events have become entrenched as units and subsequently undergo abstraction and schematization, a language learner will be able to use these units, be they single lexical items, phrases or sentences, in their own communication. This cyclical pattern substantiates the basic assertion of usage-based approaches to language acquisition, which is that language is learned through meaningful use (Langacker 2000, 2009).

1.2: A Dynamic Systems Theory approach to language acquisition

The second approach to language acquisition, which has many connections with the DUB approach, is the application of Dynamic Systems Theory (DST) to language development. In this approach language is seen as a complex, dynamic system which consists of completely interconnected subsystems. This system is in constant flux, with development based on initial conditions and growth conditions, constant variability whereby the system only temporarily settles into states of lower and higher states of variability, and for which there is no a fixed end-state (de Bot, Lowie & Verspoor 2007). While this study is not a longitudinal investigation of language development, DST does have several characteristics which are relevant for this study. The most important insight into the language systems proposed by DST is the complex and interconnected nature of the system and its subsystems. This means that, by its very nature, there is no strict division between the lexical and grammatical subsystems, or any other subsystems (de Bot, Lowie & Verspoor 2007). Accordingly, the most important implications of the principles and insights underlying DST for this study are that language development is dynamic in nature and that the separation of the lexicon from other language subsystems is necessary for studying it, but in essence artificial.
Chapter 2: Factors in vocabulary acquisition

Introduction

While the distinction between the different aspects of language may be seen as artificial from the perspective of usage-based and dynamic approaches to language acquisition, the lexicon will nonetheless be separated from the language system for this investigation. In order to investigate the acquisition of vocabulary from video games, we must first determine the nature of the developing foreign language lexicon. The most important aspects of this are the distinctions between receptive and productive mastery, and between single lexical items and chunks. Additionally, we must investigate which factors play a role in the acquisition process, the most important of which are frequency, context, and learning resources, and the difference between intentional and incidental learning.

2.1: The nature of the lexicon

The lexicon is a large and complex system of vocabulary in different languages, which has seen a substantial amount of investigation in past decades. For the purposes of this study, two aspects of the lexicon are particularly important: the distinction between receptive and productive vocabulary, and the question to what extent vocabulary is acquired as individual lexical items or chunks.

2.1.1: Receptive and productive vocabulary

In investigating the lexicon of language learners, a distinction is usually made between receptive and productive lexicon (Schmitt 2010). The distinction between the two is generally not seen as a strict dichotomy but rather a multi-faceted assembly of elements in a continuum, including factors such as partial word knowledge and a gradation between recognition, recall, and production of different aspects (e.g. phonological, orthographical, etc.) of lexical items (Webb 2008a, Schmitt 2010). Logically, this would exclude any productive aspects such as pronunciation and pragmatic information, even though this separation in the process of acquiring lexical items and chunks may be artificial (see section 2.2.2). Additionally, the focus on only one side of the lexical spectrum in this
regard means that while receptive vocabulary size may be indicative of language proficiency in general (Webb 2008a), it does not yield a full picture of overall vocabulary knowledge (Schmitt 2010).

Whichever interpretation of the receptive versus productive vocabulary distinction we use, previous research shows quite clearly that learners’ receptive vocabulary is always larger than their productive vocabulary (see Webb 2008a for an overview). While this would seem to imply that receptive knowledge is generally acquired before productive knowledge, this need not necessarily be the case. Learners may also acquire certain productive aspects of items (e.g. pronunciation, pragmatics) before it can be used with complete understanding (Webb 2008a).

2.1.2: Single lexical items and chunks

Another important aspect of the lexicon is to what extent vocabulary is acquired as chunks (sometimes labelled ‘formulas’, among many other interchangeable terms) as opposed to single lexical items. The concept ‘chunk’ has been defined in many ways (see Schmitt 2010:119 for some examples). For this study, I will adopt the definition that chunks are “[...] a word or word string, whether incomplete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any subparts it may have.” (Wray 2008:12). This definition means that chunks are seen as being stored and processed holistically as a pre-fabricated lexical sequence (Schmitt 2010). The concept of ‘chunk’ can then include such structures as idioms, sayings, strong collocations, or conventionalized discourse functions (e.g. politely declining something by saying “no thank you”). The question how these chunks are acquired is important because they make up a large part of the total contents of natural language use (Schmitt 2010). In accordance with Schmitt’s recommendations (2010), we will therefore investigate the development of the lexicon both in terms of single and multi-word (i.e. chunks) lexical items. Please note that the term ‘vocabulary’ is used to apply to both individual lexical items and chunks, and the latter two terms are only used if it is necessary to make a distinction between them.

2.1.3: Incidental and intentional language learning

In addition to the nature of the vocabulary as a system, another issue that needs to be addressed briefly in this study is the way learners may come into contact with language. Particularly important in this regard is the distinction between incidental and intentional vocabulary learning. The term ‘incidental learning’ is defined by Schmitt (2010) as “[...] learning which accrues as a by-product of language usage, without the intended purpose of learning a particular linguistic feature.” (p. 29). This is opposed to ‘intentional’ learning, where learning occurs with the intention of learning one or more linguistic feature(s). This concept of incidental learning has been researched extensively in relation to the actual and potential gains of L2 reading for pleasure by learners (see Webb 2008b or Schmitt 2010 for an
A substantial number of studies have found incidental learning to be possible depending on several factors (e.g. see Eckerth & Tavakoli 2012 for an up-to-date overview), several of which will be discussed below. In this study, the concept of incidental learning is applied to target language exposure and usage in video game playing. Therefore, the discussions of factors influencing vocabulary acquisition presented below will mostly be concerned with incidental language learning.

2.2: Frequency in language acquisition

The first of the many factors influencing the incidental acquisition of a foreign language is frequency. Frequency of occurrence and use is seen by many researchers as being one, if not *the* most important factor in the acquisition of lexical items and chunks. This factor is so important because it is significant in all aspects of vocabulary acquisition and use (Schmitt 2010). This includes how often lexical items and bundles are met by learners in input, and how quickly and accurately items are processed as well as produced (Ellis 2002). Several aspects of the role of frequency in vocabulary acquisition are particularly important for this study, namely the order of acquisition, the extent to which the different aspects of knowledge of vocabulary are acquired, the possibility of an acquisition threshold, and the dynamics of retention and attrition of vocabulary depending on frequency.

2.2.1: Order of acquisition and frequency

Starting with input side of the acquisition order, the argument is based on the logical assumption that more frequent lexical items and chunks are met more often by learners in natural or naturalistic input than lower frequency ones, although there may be some differences between different modes and purposes of language use (Schmitt 2010). This assumption is important when considering the fact that research has found that every experience of an event, be it a usage language or something else, leaves a trace in the memory of the person experiencing it (Langacker 2009). This claim is at the centre of the process of linguistic abstraction as proposed by Langacker (2009), with patterns of neural activation, for example when encountering a lexical item or chunk, strengthening synaptic connections in the brain. Further neuro-linguistic evidence comes from Tremblay et al. (2008), who found support for the trace retention claim from ERP (Event Related Potentials) studies. Consequently, several studies have found convincing evidence that more frequent lexical items are generally learned before less frequent ones, in both the L1 and L2 (see Schmitt 2010 for an overview).

2.2.2: Incremental acquisition and frequency

While the aforementioned evidence supports a frequency dependent order of acquisition of vocabulary, it does not by itself provide much information about the extent to which the different
aspects of vocabulary knowledge are acquired based on frequency of occurrence and use. The process of vocabulary acquisition is widely seen by researchers as being incremental, with the learning of lexical items moving in small steps from partial to complete (Eckerth & Tavakoli 2012). This is supported by several studies in incidental vocabulary learning. Waring and Takaki (2003) found a larger increase in the acquisition and retention of vocabulary in terms of word recognition than in terms of word recall knowledge. Additionally, Brown, Waring, and Donkaewbua (2008) found that incidental learning can take place in terms of word form recognition and the recognition of word meaning, but that far less was learned in terms of being able use the words in production. These findings provide evidence in favour of the claim that, because of its incremental nature, receptive aspects of vocabulary (e.g. form or meaning recognition and meaning recall) are likely, but not necessarily, acquired before productive aspects of the same items.

2.2.3: Acquisition threshold

Having established the importance of frequency in language acquisition, it is also important to know approximately how many exposures are necessary for incidental language learning to be possible, beyond the simple assertion that ‘more is better’.

Previous research into the incidental learning of vocabulary from reading (i.e. learning words from reading which is for pleasure and not for the express purpose of learning a language) has given some insight into how many occurrences of vocabulary are necessary. However, there appears to be no fixed number of exposures for insuring acquisition (Schmitt 2010). One study found (very limited) learning effects for intermediate learners with as few as 2, 4 and 6 exposures (Rott 1999), while Pigada and Schmitt (2006) found that the learning of meaning(s) of some words was not assured even after more than twenty exposures. The most likely average learning threshold that emerges from the many different studies investigating the subject, seems to be that between 8 and 10 exposures in reading have a reasonable chance of being acquired incidentally (see Schmitt 2010 for an overview).

With this in mind, we must also establish how much input would be necessary for enough repetitions of vocabulary to occur to facilitate incidental learning. Research into the role of extensive reading in L2 vocabulary acquisition has shown that while learners may at most acquire only a handful of words per large body of text (e.g. a novel), learners may be able to acquire the (partial) meanings of these words (Laufer & Rosovski-Roitblat 2011). This may be one reason why extensive reading in the L2 solely for the purpose of entertainment has been shown to be (to a limited extent) beneficial for L2 learners, even though corpus analysis has shown that low frequency words may not be encountered frequently enough to be acquired (Cobb 2007). However, a proverbial ‘flood’ of reading, or L2 input, could conceivably provide enough repetitions for acquisition (Laufer & Rosovski-Roitblat 2011), effectively making the acquisition process one spanning months or years.
2.2.4: Frequency and retention

Just as frequency has an effect on the acquisition of a language, so does it have an effect on the way it is forgotten. The attrition of a language is seen as a natural part of language development and has been investigated extensively in the past decades (Schmitt 2010). One aspect of this research has found evidence for a difference between the attrition of receptive and productive vocabulary. In line with past research based on the ‘dormant language’ hypothesis in language attrition studies, attrition is seen as a loss of lexical access rather than complete disappearance from memory (Bardovi-Harlig & Stringer 2010). The extent of this loss has been shown to affect productive vocabulary more than receptive vocabulary (Schmitt 2010), which could be seen as a gradual loss (of access) of certain productive aspects on the continuum between form recognition and full productive use (see section 2.2.2).

2.2.5: Frequency and chunks

While the preceding discussion focused mainly on individual lexical items, it is also important to determine the effect of frequency on the (incidental) acquisition of multi-word chunks. Starting with the simplest explanation, research has found that frequency effects for chunks are basically similar to those for single lexical items, in the sense that more exposures facilitate better retention (Tremblay et al. 2008). Similarly, chunks may also be acquired incrementally, being partially known until they have been met frequently enough to become completely acquired (Schmitt 2010). Based on the usage-based perspective on acquisition, chunks may function as more or less indivisible ‘units’ until commonalities are reinforced enough for its constituents to be individually learned (Langacker 2009). Additionally, research has found some evidence for the holistic storage of chunks in the increased processing speeds observed for chunks dependent on frequency, with lower frequency ones being processed more slowly than higher frequency ones (Arnon & Snider 2010).

2.3: The role of context in language acquisition

A second significant factor in the process of language acquisition is the context in which language occurs. With regard to the subject of this investigation, the concept of ‘context’ can be construed in several different ways, namely as being the situation in which language is learned, consisting of images, actions and experiences, and the purely linguistic context, in terms of the text and audio present in games.
2.3.1: Situated cognition and learning

The first and broadest context-related factor playing a role in vocabulary acquisition is one which affects all learning. Situated cognition is a significant factor because it affects all human learning processes, be they language acquisition as seen from a dynamic language systems perspective (de Bot, Lowie & Verspoor 2007) or other areas of learning. The underlying claim in this regard is that all learning is situated in one or more contexts because “[...]knowledge is contextualized in physical and sociocultural spaces, [...] inextricably bound to the intentional action of actors within these spaces, [who] shape and transmit knowledge by means of narrative structures.” (Neville 2010:449).

Extending this to language acquisition, we see that in situated language learning there is more to language context than purely linguistic sources alone. Situational elements such as images, sounds and actions may all be present in the learning situation, and thus play a role in that learning process (Gee 2007). One example of this is how according to situated cognition, humans organize and classify experiences according to narratives, and learners can therefore semiotically encode these experiences in a narrative (Neville 2010). Another contextual element in situated cognition is embodied cognition. This refers to a learner’s ability to experience actions performed in a specific situation, such as actions in a virtual word, as contributing to the situation they occur in (Gee 2007). This will be discussed in more detail in the next chapter on video games and language learning.

2.3.2: Linguistic context and comprehensibility

While understanding that the concept of ‘context’ in language learning is more than language alone, the purely linguistic context in which language occurs is nonetheless important. In order to acquire meanings of lexical items or chunks from the linguistic contexts (i.e. sentences, phrases, etc.) in which they occur, learners can use a process known as lexical inferencing. This is the process whereby learners make an informed guess about the meaning of lexical items in their textual context, using the co-text, and their existing knowledge about the language and the world in general (Schmitt 2010). Several studies have shown that this strategy is frequently employed by learners and that, while having low success rates, can result in the learning of the (partial) meanings of vocabulary (see Schmitt 2010 for an overview).

However, two factors have been found to negatively impact a learner’s inferencing ability. The first is proficiency in the target language, with lower proficiency learners having a lower success rate in inferencing that higher proficiency learners (Liu & Nation 1985). The second is the comprehensibility of the input from which learners infer the meaning of items. Learners have been found to make more successful inferences from texts with a higher percentage of known words than from texts with a lower percentage of known words (Liu & Nation 1985). This last factor is especially important, because it implies that immersion in a FL environment where a learner only
understands a very limited percentage of words can logically be assumed to be completely incomprehensible, especially if there is no assistance of any kind (i.e. linguistic or extra-linguistic) available to make it more comprehensible.

2.3.3: Contextual salience

An important element in the acquisition of vocabulary learning from context is what is called the ‘salience’ of a lexical item or chunk. The concept of salience has been defined as “[t]he perceived strength of stimuli [...]” (Ellis & Ferreira-Junior 2009:331). It is beyond the scope of this study to examine the ways in which ‘salience’ has been conceptualized, but it does examine certain data (see chapter 4) in order to find those test items which drew the attention of the participants in the study. Consequently, salience is seen as being somewhat personal in nature, with the strength of a stimulus being dependent on a learners’ perception.

2.4: Resources for language acquisition

Finally, in addition to the factors discussed above, there are several extra-linguistic factors which have been shown to impact the process of acquiring a foreign language. These constitute the conditions needed for acquisition to occur, which can be grouped into internal and external learning resources (de Bot, Lowie & Verspoor 2007).

2.4.1: Internal learning resources: attention, involvement and engagement

Internal language learning resources are generally considered to be those related to a learner’s mental capabilities, including such factors as conceptual resources, motivation and learning capacities. They are inherently finite, due to a learner’s limited memory capacity and motivation to learn (de Bot, Lowie & Verspoor 2007). The most important of these for this study is attention, because of its role in the incidental acquisition of language. Considering the extreme unlikelihood of ‘subconscious’ or completely ‘unaware’ acquisition of vocabulary, we have to presume that some level of attention, however minimal, is needed in order for elements of FL input to become intake and then be acquired by a learner (Schmidt 2010).

The progression from input to intake to acquisition starts with a learner’s noticing of new language. Although this central claim to Schmidt’s noticing hypothesis has been disputed on several points (Schmidt 2010), there appears to be enough evidence to support it. If and when a language form is noticed, the level of attention that is given to its different aspects may determine how well it is acquired and retained. Schmitt (2010) made the straightforward claim that more of what he calls ‘engagement’ with a linguistic unit leads to better acquisition and retention. Laufer and Hulstijn
(2001) framed this proposition in the form of the Involvement Load Hypothesis (ILH), which proposes a combination of the need for comprehension of an item, searching for its meaning, and reflecting on this meaning in its context as the three factors influencing the amount of attention given to an item. Although this combination of factors may not cover all aspects of engagement with and attention to new language forms (Schmitt 2010), it has nonetheless been supported in additional research (see Keating 2008).

This discussion of attention and task-induced involvement is of particular importance to the process of incidental language acquisition. As discussed above, the act of performing a task such as reading for pleasure and the potential gains in language resulting from this task rest on the learner’s noticing of, attention to, and involvement with new language forms.

2.4.2: External learning resources: modality and input enhancement

In addition to internal resources, the effectiveness of language learning also depends on many external resources not related to an individual learner’s mental capabilities (de Bot, Lowie & Verspoor 2007). The most important of these for this study is the use of material to support language learning, specifically the use of subtitles and closed captions in visual media. This issue of input modality was investigated by Sydorenko (2010), who found that the addition of target language captions to a target language video input improved word form recognition and meaning acquisition, whilst on the other hand this detracted from aural word recognition. Additionally, she found that learners pay most attention to the target language captions in this situation, followed by video and audio. Finally, she also found that most acquisition was achieved by association between words and visual images. Several other studies have also shown that enhanced audio-visual input through captions may be beneficial for incidental vocabulary acquisition (see Perez & Desmet 2012). However, as Perez and Desmet (2012) point out, the inconsistent methodologies used in this area of research and the variable nature of the different lexical knowledge aspects that are acquired by learners make it very difficult to generalize these findings into one coherent assessment of the effectiveness of subtitles and captions.

2.5: Dynamic language development

Finally, it is necessary to consider a dynamic view of vocabulary learning as proposed by DST. Several aspects of this overarching theory may be instructive for this investigation. In particular, it is the fundamental principle of constant variability in development which may have implications for vocabulary acquisition from video games.

As was briefly discussed in the first chapter, the dynamic view of development is based on the concept of fluctuating degrees of variability and stability in the developmental process. Evidence for this concept has usually been found by closely examining the development of acquisition and use of
certain language elements in longitudinal studies. These studies have yielded two important observations that may be important for this study.

Firstly, due to the interconnected nature of different aspects of language, the development of these aspects have been shown to be interrelated at different levels. These relations may be either supportive, meaning that two or more structures are used more at the same time, or competitive, meaning that one usually more advanced form is used more while another less advanced form is used less (Spoelman & Verspoor 2010). Observations of these relations have been made in several different studies, for example in relation to L2 Finnish morphological accuracy (Spoelman & Verspoor 2010), syntactic complexity in L2 English (Verspoor & van Dijk 2011), and the development of academic L2 English in terms of different passive and productive language skills (Caspi & Lowie 2010).

Secondly, the observation of interconnected development of different language structures, together with the assertion of DST that there is no end-state of development, means that there is always variability in development. Although it is possible for relative stability, in the form of so-called ‘attractor states’, to occur at some points in development, variability never completely disappears (de Bot, Lowie and Verspoor 2007). This has been observed in several studies (see Verspoor & van Dijk 2013 for a brief overview). What appears particularly informative for this study is the way in which this affects the developmental trajectories of aspects of language at different levels. At the level of the lexical system, research by Caspi and Lowie (2010) showed complex interrelations between passive (active recall and recognition) and productive (controlled and free production) knowledge in a student’s academic L2 English vocabulary development. The variability of this development was most clearly visible in the shifts between supportive and competitive relationships of these skills. At the level of specific constructions of L2 development in particular language domains, significant variability was found in Spoelman and Verspoor’s (2010) study on the development of Finnish morphology, and in van Dijk, Verspoor and Lowie’s (2011) findings of high variability and consequent overgeneralisation of new L2 forms by learners.
Chapter 3: Video games as language learning environments

Introduction

Having established the most important factors in the incidental acquisition of foreign language vocabulary, this chapter focuses on how factors such as repetition, context and learning resources can be found or implemented in video games. First, I briefly discuss how language learning can be achieved through video games which are expressly created for the purpose of language learning. Next, I discuss the most important aspects of non-educational video games that can be conducive or supporting for language learning, based on previous research. Finally, I summarize the findings from previous studies and relate them to the purpose of the present study.

3.1: Terminology and typology: educational versus non-educational video games

First, for the sake of clarity, it is useful to examine an educational video game in order to be able to formulate a distinction between educational and entertainment-oriented video games. However, making this distinction is not straightforward because video games designed purely for teaching something to its players can and must also be entertaining to a certain extent. On the other hand, video games designed solely for entertainment purposes may, and in most cases do, have some educational properties as well. This has been investigated by a growing number of studies and books (see for example Prensky 2006, Gee 2007, McGonigal 2011). Nonetheless, in order to investigate the properties of educational video games for FL or L2 learning, I have chosen to partially use Klopfer et al.’s (2009) definition of what they call ‘learning games’ as being those games that “[...] target the acquisition of knowledge as its own end [...]” (21). One example of such a game is Ubisoft’s My Word Coach (2008), which was studied (and partially developed) by Tom Cobb (Cobb & Horst 2011). It is intended to facilitate rapid L2 vocabulary expansion, both receptive and productive, through several word games with both focus on form and meaning. The core principle underlying the game is repetition, with lexical items being recycled based on how often they are encountered and correctly used in a particular game.

The other group of video games with which this study is concerned has usually been named ‘commercial-off-the-shelf’ (COTS) games, which are “[...] not specifically tailored to L2 learning or teaching [...]” (Cornille, Thorne, Desmet 2012:246). However, it is specifically this category which has increasingly been the subject of study. The effects of playing COTS games has been investigated (more or less formally) in relation to many different learning processes, skills and school subjects. Young et al. (2012) provide an extensive overview of the use of video games for school subjects such as mathematics, science, history, and language learning. Language learning in particular seems to stand out in much of the literature because of several properties inherent in most video games which
can facilitate it, possibly more so than other subjects. These include possible FL interaction between and among native speakers (NSs) and language learners (LLs), repeated exposure to FL vocabulary, and contextualized language use. On the other hand, there are also several factors which may diminish or even completely prohibit the possibility of language learning, the most important of which are incomprehensible FL input and insufficient attention to language due to overwhelming cognitive load.

3.2 Factors facilitating incidental language acquisition through COTS video games

3.2.1: Safe and meaningful interaction in multi-player gaming

The first major attribute of video games which enables them to provide a basis for incidental FL vocabulary acquisition is interaction between players, either through a shared online or offline gaming experience. Most research into FL acquisition from COTS games to date has focused on games as interactive media through which and about which meaningful communication between players can occur. This interaction can occur between NSs and learners of the target language, between learners, online (through text and voice chat) or offline (local multi-player), and whilst playing multi- or single player games.

Interestingly, in the past few years most of this research has focused on the potential of so-called massively multi-player online role playing games (MMORPGs) for language learning, specifically World of Warcraft (WoW) (Blizzard Entertainment 2004), which is the largest of the genre outside Asia with 11 million players worldwide (Rama et al. 2012). The collaboration-oriented nature of this genre of games necessitates nearly constant communication between players, be they native or non-native speakers of a language. This has made MMORPGs a good genre for studying FL learning through interaction.

Perhaps the most important LL conducive aspect found by researchers, is the way in which these types of games are able provide learners with a ‘safe’ environment for meaningful use of the target language. Research suggests lower inhibitions for TL use (Reinders & Wattana 2011), increased risk taking through role-playing (Peterson 2010), and an overall increased level of comfort in communication as play progresses (Rankin et al. 2009). There is evidence that this could be due to anonymity in gaming (Zhao and Lai 2008), and more importantly the scaffolded and comprehensible interaction between more experienced and less experienced players, the former of which may be seen as “[...] catalysts for language socialization.” (Thorne, Black & Sykes 2009:810). This highlights the significant value for LLs to have direct contact with, and receive immediate feedback from, NSs through a game, facilitating several aspects of L2 development. Thorne (2008) describes a conversation between a Ukrainian L2 English speaker and an American native English speaker (NES) in which the Ukrainian speaker draws explicit attention to his produced English language forms for feedback such as spelling correction. Rama et al. (2012) found an improvement in a participant’s
communicative competence in an L2 Spanish WoW environment, balancing his low level of Spanish with his extensive WoW experience. Similar results were found in communication between ESLLs and NESs through MMO gaming (Rankin et al. 2009). Finally, Piirainen-Marsh & Tainio (2009) showed that interaction between two LLs in the same location sharing a single-player gaming experience (i.e. one plays, the other watches) can “[...] co-construct their understanding and enjoyment of the game[...]” (175) through repetition and imitation of language use in the game. While these findings do not have direct implications for the particular way in which COTS gaming is investigated in this study, it is important to keep in mind the potential for interaction between gamers even for the type of video game discussed here.

3.2.2: Frequent and complexly distributed recurrence of foreign language vocabulary

The second major factor enabling video games to serve as good grounds for incidental FL vocabulary acquisition is the frequent recurrence vocabulary. In chapter 2.2, we saw that a significant amount of research has been able to identify frequency as a key factor in vocabulary acquisition. In relation to video games, frequency of occurrence of FL vocabulary is important in two ways, namely in the sense of its distribution, and the frequent recurrence and reinforcement of game related vocabulary.

Starting at the textual level, the distribution of FL vocabulary is important because it determines how lexically complex a text is, and consequently how frequently the items and chunks occurring in those texts usually occur in other contexts and uses (Thorne, Fischer & Lu 2012). In recent research, Thorne, Fischer and Lu (2012) used a range of lexical and syntactic complexity measures to investigate the distribution of English vocabulary in so-called ‘quest texts’ (assignments given to the player) in World of Warcraft. They found that while the readability of the texts was concurrent with grade 7 or 8 US middle school levels, the texts contained relatively high numbers of low-frequency items, complex syntactic structures, and a high level of lexical diversity. Therefore, this explorative study shows that language in video games can be simultaneously readable, and highly sophisticated, lexically diverse and syntactically complex. However, the researchers also address several significant problems which make it difficult to generalize these findings to other video games. The language content may differ considerably between individual games and between genres, as deHaan (2005b) proposes by making the generalization that RPGs are typically more story-oriented and therefore linguistically complex than sports games. Additionally, Thorne, Fischer and Lu (2012) raise the important point of the impact of genre on the way language is used, with games including dialogue and text mainly addressed to the player (i.e. in second person perspective), possibly showing lower lexical sophistication than those using mainly third person perspective for expository or explanatory functions.

Ultimately, the problem for the current study is the lack of previous research into this aspect of video games, yielding no generalizable findings to support useful assumptions about the linguistic complexity of the game used in this study. Additionally, complexity analyses such as those performed
by Thorne, Fischer and Lu (2012), were not applied to the corpus created for this study due to time constraints. However, the author would like to propose that narrative-driven video games, independent of genre, may be necessarily linguistically complex due to story-writers’ aims to make dialogues and texts appear as natural as possible. This speculative proposition means that we may be able to safely assume that narrative-driven first-person-shooter (FPS) games like *Half Life 2*, used in this study, are more linguistically complex than sports games, which typically do not have the same depth of story, dialogue, and character development as narrative-driven FPSs and RPGs.

While the distribution of FL vocabulary in the story told by certain games is of importance for how many times target items and chunks are met in input, there are several other sources of input unique to video games which may contribute to the acquisition of FL vocabulary by players. As was discussed in section 2.2.3, FL vocabulary must be met multiple times in input before it can be acquired. In this regard, narrative-driven games such as the one used in this investigation have two relatively distinct sides. On the one hand, as was discussed above, the story is presented through texts, dialogues and monologues that may contain a similar distribution of vocabulary in terms of uses of low-frequency words and complex syntax. On the other hand, almost all games utilize certain mechanisms to make the game playable through intuitive menus, in-game heads-up-displays (HUDs), and sometimes routinized non-playing-character (NPC) teammate communication. In this regard, most story-driven games may be viewed as partly similar to sports games in terms of the frequently repeated and situated use of (relatively high frequency) vocabulary present in the game’s monologues, menus and HUD, among other sources. The most concrete example of the effect of this on learners is deHaan’s (2005a) study, showing that repeated exposures to baseball-related Japanese FL vocabulary in a game’s menus, HUD, texts and sports commentators’ formulaic announcements can lead to acquisition of that vocabulary. In this way, video games, more so than other forms of entertainment media, may offer certain unique sources of repeated FL vocabulary in reinforcing contexts, which is what will be discussed in the next section.

### 3.2.3: Situated cognition and lexical infeference in COTS games

The second major factor enabling incidental FL acquisition from video games is the situated and contextualized use of language throughout video games, specifically the unique way in which games can use situated and embodied cognition to create meaning, and the way in which the context (in both textual and audio sense) can be used to provide the meaning(s) of words.

Starting with situated and embodied cognition, it has already been discussed how these two phenomena can help construct meaning of language in a broader, non-linguistic context,. Video games offer unique opportunities for embodied and situated learning, of language or any other skill, which are not found in other entertainment media. We have already seen the definition of knowledge as contextualized and situated proposed by Neville (2010) based on his research of the literature on this
topic. Similar to Neville’s (2010) formulation of the term ‘knowledge negotiation’, Gee (2007) makes the claim that meanings in video games are always situation specific and ‘assembled’ by players through the images, texts and embodied actions in the game world. Therefore, he claims, “[t]o make sense of [words] [a player] must fit them into the emerging plot and virtual world [a player] [is] discovering and helping to build.” (Gee 2007:83).

One anecdotal example of situated cognition is the understanding of the word ‘headcrab’ by a gamer who has played *Half Life 2*, which will most likely differ greatly from those who have not played the game. The latter group of people most likely construct the meaning of the word ‘headcrab’ by assembling the images, sounds and their experiences fighting them into a coherent definition of the word. Without this situated and embodied information, many people may be able to combine the images of a crab and a head, but would probably not be able to imagine a ‘headcrab’ to be a small, four-legged alien creature which attaches itself to the heads of NPCs or the player, turning them into zombies. Furthermore, they would never have the embodied experience of repeatedly nearly being ‘killed’ by an extremely dangerous ‘poisonous headcrab’, which hurls itself at the player from a dark corner with a blood-chilling screech. The insights gained from this view of (language) learning for this study are that, as meaning is situated and embodied, learners may acquire FL vocabulary from video games by assembling information they find inside (and possibly also outside) of the video game environment.

Some indirect evidence for this building of meaning from different sources may be seen to come from Sydorenko’s (2010) research, in which she found that learner’s watching a captioned video frequently used the strategy of associating (target language) TL words with images in the video. In addition to facilitating this connection between images and words, the provision of captions may also enable some learners, especially higher-level learners, to use lexical inferencing (discussed in section 2.3.2) in order to determine the meaning of FL vocabulary encountered in the game. Furthermore, many games include large bodies of text which including both relatively simple and complex language (Thorne, Fischer, Lu 2012) which might enable lexical inferencing.

### 3.3 Factors inhibiting incidental language acquisition through COTS games

While the above sections discuss several properties of non-educational video games which create the potential incidental FL acquisition, there are also several factors which may decrease or completely remove this potential. The most important of these factors are the possible incomprehensibility of the FL input the game provides and the possibly extraneous cognitive load placed on players by the game’s interactivity.
3.3.1: Comprehensibility of input

Although regular video game players may be provided with a ‘flood’ of TL input, potentially facilitating incidental acquisition of TL forms (Laufer & Rozovski-Roitblat 2011), this input does need to be comprehensible for it to be adequately processed by a language learner. Different types of video games may be able to provide learners with several ways to make input comprehensible, although it remains a problematic issue.

As was mentioned earlier, studies suggest that comprehensibility of FL multi-player games can be enhanced by interaction with NSs of the FL. Rankin et al. (2009) found that ESL students playing an MMORPG in English frequently sought help from NESs in order to make sense of the game’s mechanics and contents, which even went so far as the learners exhibiting ‘follow the leader’ behaviour with respect to English NSs. Furthermore, Piirainen-Marsh and Tainio (2009) found that learners sharing a TL single player experience using mechanisms such as ‘other-repetition’ to collaboratively gain understanding of the game. These studies provide some initial evidence for the possibility of players of different language backgrounds collaboratively making sense of the language contents of a foreign language video game, either online or offline.

In contrast, single player games which are played alone and in isolation from possible help with the TL, provide a more problematic situation in terms of TL comprehensibility. The only possible way to mitigate this problem seems to be enhancing single player games with content in the target language or players’ native languages. This may be achieved through use of subtitles and captions already provided by the game for hearing-impaired players, or by modifying a commercially available game. Purushotma (2005) modified the German language version of the most popular game in history, The Sims, to include translations of object descriptions and word glosses in English, essentially providing a multilingual environment, which could improve its learning potential for English speaking learners of L2 German. Additionally, as previously discussed in section 2.4.2, captions and subtitles in the TL may be included in media to help learners make sense of the audio input they receive. The beneficial effects of including not translated captions but captions in the same language as the audio input has been shown to be beneficial in previous research, although this was all conducted on video instead of video games (see Perez & Desmet 2012 for an overview). This may be problematic due to the interactive nature of video games (as opposed to video), which could divert attention away from captions to gameplay or contribute to extraneous cognitive load, which will be discussed in more detail in the next section.

Consequently, despite these possibilities of enhancing FL games, immersion in that language in a gaming environment without the aid of fellow players or additional LL resources (e.g. dictionaries, grammar references) does have the potential to make the language a thoroughly incomprehensible environment for beginning language learners, even with captions. Rankin, Gold, and Gooch (2006) found support for this in one ‘high-level beginner’ ESL participant’s difficulty with
adapting to the game they used. This is why they suggest that gaming for language learning is only beneficial for intermediate and advanced learners. Ultimately, although learners of different levels have different learning needs, a non-educational video game which has not been enhanced in any way to facilitate language learning remains a medium which “[...] may lack the flexibility to support the needs of varying levels of ESL learners [...]” (p. 4), or indeed of any language learners.

3.3.2 Attention to language and cognitive load

Another factor limiting the potential of games for language learning are the natural limits of learners’ cognitive resources and the overwhelming effect game playing may have. Investigations into this problem have focused on the difficulties faced by game players, especially beginners, with gaining control of the game and simultaneously reserving some cognitive resources for attending to language in the gaming environment.

Although Purushotma (2005) showed that some interactive enhancement of game interfaces may be conducive to incidental FL acquisition (see 3.3.1), two other studies have shown that the many interactive elements of game play can be problematic. Rankin, Gold, and Gooch’s (2006) findings indicate that the steep learning curve associated with mastering navigational control in a video game could have resulted in cognitive overload for those participants who also had a relatively low (i.e. ‘high-level beginner’) ESL proficiency. The participant in question needed to divide his/her cognitive resources between too many new experiences, namely (1) mastering control of the game, (2) paying attention to the images, sounds and textual information in the game, and (3) making sense of this information which could have been incomprehensible to a certain extent due to the participants’ low level of proficiency. A somewhat similar effect of interactivity was found by deHaan, Reed, and Kuwada (2010), who specifically tested the differences in EFL vocabulary recall by one group of active game players and one group of passive observers. The researchers found that both groups did recall some vocabulary, but that the participants who had only watched the game being played by a participant from the other group in a different room, recalled significantly more target vocabulary than those participants who had to play the game. These findings were supported in participants’ own perception of the difficulty of the language they encountered in the game, with passive observers finding the language easier to comprehend. Additionally, active players reported that they experienced difficulty with concentrating on the language in the game when they were forced to control the game mechanics, a finding similar to what the Japanese FLL participant in deHaan’s earlier (2005) study indicated, although he also indicated that he was able to pay more attention to language as he gained better control over the game’s mechanics and lowered its difficulty level.

In light of this initial research, it appears that extraneous cognitive load placed on a learner by a combination of controlling game mechanics, and paying attention to and processing new language encountered in the game, can inhibit incidental FL vocabulary acquisition from video games.
However, video games, especially those revolving around a complex story moving the game forward, often intersperse longer sequences of active playing with shorter so-called ‘cinematic cutscenes’ during which control is usually to some degree taken away from the player as plot unfolds through animated videos or pre-scripted in-game story sequences with text and audio. In line with DST (de Bot, Lowie & Verspoor 2007), one could make the assumption that the allocation of cognitive resources such as attention is in constant flux. I would therefore suggest the possibility that in playing (certain types of) video games, players may be able pay more attention to language at some points during play (i.e. during cut-scenes) due to the fact that, for a limited time, they do not have to attend to controlling and navigating through the game.

3.4 Statement of purpose

Considering all of the factors discussed above, it seems that single-player only video games which are driven by a complex narrative, for example Half Life 2, may be the least conducive games for acquiring FL vocabulary. While they use frequently repeated and contextually situated lexical items and chunks, they provide no possibility of interaction with NSs, except in the case of a shared offline single-player experience, they probably do not offer comprehensible FL input for any learners but those at a higher proficiency level, and may cause extraneous cognitive load leading to attentional difficulties. However, this type of video game is very popular, with games such as Mass Effect 2, Final Fantasy XIII and Red Dead Redemption being among the top 20 most sold games in the year 2010 (ESA 2011). Many players around the world play these games, which are directed at adolescent and mature audiences, in a foreign language: English. This means that many players raised with native languages spoken by relatively few people on earth, are exposed to large amounts of a foreign language, and possibly learning some of it while simply engaging in their hobbies.

The purpose of this study is therefore to find the answers to the questions (1) can gamers incidentally acquire FL vocabulary from video games and (2) what is the effect of facilitating and inhibiting factors on the incidental acquisition of FL vocabulary from video games?

In order to find answers to these questions, several hypotheses are formulated:

Hypothesis 1: Playing a COTS video game in a foreign language leads to acquisition of vocabulary from that language.

Hypothesis 2: Several factors can facilitate the incidental acquisition of foreign language vocabulary from COTS video games.

Hypothesis 2A: More high frequency FL vocabulary is incidentally acquired from COTS video games than low frequency vocabulary.
Hypothesis 2B: The context, in terms of textual, situated and embodied context, in which foreign language vocabulary in a COTS video game occurs facilitates incidental acquisition of that vocabulary.

Hypothesis 2C: Enhancing foreign language input from COTS video games by including subtitles and closed captions facilitates the incidental acquisition of vocabulary from that game.

Hypothesis 3: Several factors can inhibit the incidental acquisition of foreign language vocabulary from COTS video games.

Hypothesis 3A: Low level FL proficiency of gamers inhibits or prevents the incidental acquisition of vocabulary from a COTS video game in that foreign language.

Hypothesis 3B: Extraneous cognitive load on gamers playing a foreign language COTS video game inhibits attention to language in the game and thus inhibits incidental acquisition of vocabulary from that game.
Chapter 4: Methodology

4.1: Participants

The participants taking part in the study were seven Dutch students at a university in the Netherlands. Five participants were recruited from the applied linguistics minor program of the university, all of whom had negligible of Spanish language proficiency. Later, two more participants were recruited, one from the applied linguistics minor who had had some training in Spanish and one who was enrolled in the Romance languages bachelor program and specialized in Spanish. All participants were asked to provide information on their language skills and gaming habits, which is summarized in the table below. As far as their language skills were considered, participants were asked to rate their proficiency in the languages they spoke, an average of which can be seen in the table. Of the seven participants, two had previously received or were still receiving instruction in Spanish.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>28</td>
<td>25</td>
<td>23</td>
<td>19</td>
<td>22</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>Sex</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Study</td>
<td>Ba English Language &amp; Culture</td>
<td>Ba Dutch Language &amp; Culture</td>
<td>Ba Linguistics</td>
<td>Ba Classical Languages &amp; Cultures</td>
<td>Ba Romance Languages &amp; Culture (Spanish)</td>
<td>Ba English Language &amp; Culture</td>
<td>Ba Romance Languages &amp; Culture (French)</td>
</tr>
<tr>
<td>Dutch</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>English</td>
<td>5</td>
<td>4.3</td>
<td>3.3</td>
<td>3.3</td>
<td>4.3</td>
<td>3.7</td>
<td>5</td>
</tr>
<tr>
<td>French</td>
<td>1.7</td>
<td>3.3</td>
<td>1</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>4.3</td>
</tr>
<tr>
<td>German</td>
<td>1.7</td>
<td>2.3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>N/A</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>Mandarin: 1.3</td>
<td>Frisian: 4</td>
<td>N/A</td>
<td>Spanish: 2</td>
<td>N/A</td>
<td>Spanish: 2.6 Lithuanian: 2</td>
</tr>
<tr>
<td>Average hrs. p/w</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>FPS experience?</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Played HL2 before?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>How many times?</td>
<td>N/A</td>
<td>N/A</td>
<td>incomplete</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>N/A</td>
</tr>
</tbody>
</table>

All participants were students of different languages or linguistics and were proficient in more than one language. Concerning their gaming habits, the participants were asked how much time they spent on playing video games (average hours per week), whether they had any experience playing FPS games, and whether they had played the game *Half Life 2* before. The average time the participants spent on video games every week varied between 0 and 20 hours, and of the two participants who had played *Half Life 2* before, only one had completed it. As will be discussed in more detail in section 4.4, they completed all tests at home and had very limited interaction with the researcher.
4.2: Materials

The study used several different materials to expose the participants to the target language and to test whether this exposure led to an increase in target language vocabulary.

4.2.1: The video game

The video game used for this study was *Half Life 2* for PC, which was developed by Valve Software and released in 2004. It fits into the genre ‘first-person shooter’ (FPS), in which the player sees the action through the eyes of the story’s protagonist and navigates a virtual 3D environment, using weapons and tools to defeat enemies and accomplish goals set by the game. The game is the second instalment in the *Half Life* series. It is set in a dystopian future in which humanity has been enslaved by the rulers of a vast inter-dimensional empire referred to as ‘the Combine’. The subject plays the part of the game’s silent protagonist, named Gordon Freeman, who attempts to overthrow the Combine and their puppet ruler Dr. Breen.

The prominent role of the story and its characters driving it puts *Half Life 2*, at least in the author’s opinion, into the possible category of narrative-driven video games described in chapter 3 of this study’s background section. The game can be seen as a rich source of input in the target language. Spanish was chosen as the TL because the Dutch participants are not usually exposed to this language in their studies or recreational media. This is not the case with English, which is pervasive in many aspects of a university student’s daily life in the Netherlands. There are several ways in which the participants were exposed to the Spanish language input.

First of all, storytelling and exposition occur through so-called ‘in-game cutscenes’, during which the action is temporarily halted and navigational control is limited or completely taken away from the player. The player cannot progress until the end of the cut-scene. In *Half Life 2*, the protagonist is completely silent and therefore most of the explanation of background and story during these cutscenes is done through dialogues between NPCs or by NPCs addressing the player.
Figure 1 Showing a cutscene in which an NPC speaks to the player. The picture also clearly show the different HUD elements visible to the player, with the bottom ones always visible and the those at the top dropping down when selecting weapons.

The second source of input are so-called ‘Breencasts’, which are short video messages by the game’s antagonist, in which he addresses the people (and in one instance, the player) using large video screens in different locations in the game environment. The player always encounters these messages, but while exposure in this way is forced, the player retains navigational control of the character and can therefore ignore the message and proceed with the game.

Figure 2 Showing a ‘Breencast’ in which the game’s antagonist addresses the people of City 17 through large video screens at different locations.
The same is true of the third source of input, which are the relatively frequent shorter encounters with other NPCs in which they speak to the player or to each other. Some of these encounters are forced due to the linear nature of the game, although there are some hidden NPCs which may not be encountered. These encounters provide at most four or five sentences of input.

Fourth, there are many occasions when enemies and allies say things about a current situation. These instances of allies helping the protagonist or enemies communicating with each other usually consist of a limited number of repeated formulas. Exposure to these formulas or chunks in the target language is inevitable. However, due to the fact that their occurrence depends on factors such as how long a player takes to play through a level or how many NPC enemies and allies are in the immediate area around the player at any given time, the exact frequency of exposure is impossible to determine.
Fifth, there are also several textual sources providing target language input, namely the heads-up display (HUD), the game menus, and the closed captions. Some parts of the HUD are always on screen, namely the gauges indicating the protagonist’s health, the charge level of his combat suit and the amount of ammunition he has left for his weapon. Whenever the player changes weapons, a menu in the HUD is briefly visible detailing the player’s choices with the names of the different weapons available to him or her. Because this information is always or very frequently on screen, exposure is inevitable and highly frequent. The same is true of the primary menus needed for saving, loading and playing the game. These menus were all presented in the target language when the participants were playing the game. However, while exposure to these primary functions and menus is inevitable, all other menus (e.g. game settings, credits) can be avoided.

Finally, the participants played the game with both subtitles and closed captions in the target language. The subtitles showed a nearly perfect transcription of the dialogue between characters, while the closed captions, which are usually used by hearing impaired players, transcribe sounds and events happening around the player.

4.2.2: The corpus

In order to measure the players’ acquisition of the target language from the game, a language corpus was created, based on the audio and video recordings of a single play-through of the game made by the author of the study. He transcribed all language elements occurring in the game that were also visible in the subtitles and the closed captions. Some parts of the script were later revised and a small amount of language content, which may not have been encountered by every player, was included. The author had negligible target language proficiency at the beginning of the project.

After the script was made, a frequency corpus was created using the CLAN program developed by the CHILDES project (MacWhinney). This corpus showed the frequency of occurrence of every single lexical item in the transcript of the game. Additionally, frequently occurring chunks were identified by a near-native speaker of Spanish, who had lived and worked in Spain for several years. All of the test items used in the test described in the next section were then extracted from this corpus.

4.2.3: Testing materials

This mixed-methods study used three different measures to get information about the participants’ development in the target language vocabulary. The first two tests, a checklist questionnaire and a chunk completion test, were used to quantitatively determine whether the participants’ target language vocabulary had grown. The third test, a retrospective questionnaire, was used to determine the participants’ views of the importance of several factors in language acquisition.
4.2.3.1: Yes/no checklist test

In order to determine whether and to what extent the participants acquired FL Spanish lexical items from the video game, a computerized Yes/No checklist test (also called checklist questionnaire) was constructed. The checklist format was chosen because it has been proven to be a fast and reliable way of determining the receptive vocabulary size of language learners (Mochida & Harrington 2006).

The test consisted of a total of 100 items. 80 lexical items were taken from the game’s corpus and 20 pseudo-words were created by slight alteration of real Spanish words with no equivalents in the corpus. The real words from the game corpus were subdivided into four groups, based on their frequency of occurrence. The items in the first group all either occurred highly frequently in the game (i.e. more than 10 times) or were expected to be known to the participants prior to playing the game due to their cognate nature (e.g. Spanish “doctor” and Dutch “dokter”) or due to them being perceived as ‘common knowledge’ (e.g. ‘hola’). The second group consisted of items which occurred more frequently than the probable acquisition threshold of between 8 and 10 occurrences (see section 2.2.3) or which occurred in highly frequent NPC communications (explained in section 4.2) and were therefore almost certainly encountered more times than the same acquisition threshold. The third group of items consisted of words which occurred between 5 and 8 times in the corpus, around the acquisition threshold discussed in section 2.2.3. The fourth group of items consisted of low-frequency items which only occurred once in the entire corpus. Finally, the fifth group of items consisted of pseudo-words. These were slightly altered words not occurring in the corpus and having no meaning in Spanish. All pseudo-word items were checked by a near-native speaker of Spanish to ensure that they did not have any meaning in Spanish.

4.2.3.2: Chunk completion test

In addition to the Y/N test, a chunk completion test was constructed to measure the participants’ receptive knowledge of Spanish chunks occurring in the corpus. This test required participants to complete the last word of a chunk taken from the game corpus, of which only the first letter was provided, by choosing from four options. These options consisted of one correct answer and three distractors, which were other Spanish words. Again, a near-native speaker of Spanish checked the different options to ensure that only one was part of a commonly used chunk.

4.2.3.3: Retrospective questionnaire

Finally, all but one of the participants were also asked to complete a reflection assignment. This assignment consisted of a number of open questions concerning the participants’ experiences with playing the game. Specifically, they were asked to self-assess the impact of factors such as context, frequency and the use of subtitles on their acquisition of FL Spanish vocabulary. Because of the small
sample size and complicated nature of this study, this questionnaire was used to test all hypotheses posited in chapter 3.4, except for hypotheses 1 and 2A.

4.2.4: Scoring quantitative data

The measures were coded in different ways. First, the yes/no test was scored by assigning a value of 1 for every ‘yes’ answer and a value of 0 for every ‘no’ answer. Second, in a similar way, every correct completion of a chunk in the second test was assigned a score of 1 and every incorrect completion was assigned a 0. Next, the results of the Y/N test were recalculated, once for the total scores per test and once for each frequency category, using a calculation method developed by Huibregtse et al. (2002). This method corrects for sophisticated guessing and participants’ response style by inserting the proportion of hits (‘Yes’ answers to real words, labelled $h$) and false alarms (‘Yes’ answers to non-words, labelled $f$) into the equation presented in figure 6. This produces a figure called the ‘Index of Signal Detection’.

$$I_{SDT} = 1 - \frac{4h(1-f)-2(h-f)(1+h-f)}{4h(1-f)-(h-f)(1+h-f)}$$

Figure 6 Index of Signal Detection score calculation (Huibregtse et al. 2002). With $h = \text{hit rate (number of hits divided by total number of real words)}$ and $f = \text{false alarm rate (number of false alarms divided by total number of non-words)}$

The resulting figure is more accurate than a simple tally of the number or proportion of hits and false alarms, because the calculation method uses the principles of signal detection theory. According to these principles, participants perceive both options on the test (i.e. ‘Yes’ and ‘No’) to have a certain validity, and choose the most plausible option. The first calculation method based on this theory was later adapted to include such factors as sophisticated guessing and the personal response styles of the participants (Huibregtse et al. 2002). Thus, the $I_{SDT}$ score calculation developed by Huibregtse et al. (2002) incorporates many factors influencing participants’ testing behaviour and is therefore considered to be the most appropriate one for this study.

4.2.5: Coding qualitative retrospective data

A qualitative analysis was performed on the retrospective analyses completed by all but one of the participants. This was accomplished by close reading of the participants’ writings and coding instances in which they referred to the issues under investigation in this study, along with whether these issues were seen as having a positive, neutral or negative effect on the incidental acquisition of FL vocabulary. For example, if a participant wrote that he or she was not able to pay attention to the language due to extraneous cognitive load, this would be coded as ‘CL -’. On the other hand, an
instance in which a subject indicated being able to employ lexical inferencing successfully was coded as ‘Ctx-inf +’.

After coding, the number of each code was tallied for every participant in order to gain an indication of any general trends in the participants’ perceptions of their vocabulary acquisition from the video game. Of course, this qualitative method relies on the honesty and completeness of the participants’ self-assessments. Therefore, the results of this part of the investigation may serve to indicate what the participants consider to be important, acting in good faith with regard to their responses.

4.3: Procedures

The testing and playing period for this study spanned more than six weeks in total. In order to approximate the environment in which most gamers play a game, all testing and gaming were done at home, with the tests being available to the participants at any time via the internet. Before the participants were allowed to start playing the game, they completed the first Y/N and chunk completion test online through the website surveymonkey.com. They were also given instructions on how to play the game. The participants were not allowed to use any supporting materials and were instructed to play through the whole game once with Spanish audio and subtitles. After they had completed the pre-test, the participants were given four weeks to complete the game. This was thought to give them enough time to complete it once. The participants were free to choose the difficulty level of the game, when they played and how long each play session lasted. Immediately after completing the game, the participants completed the second test. This meant that not all participants completed the tests at the same time. Precisely two weeks after completing the second test, the participants completed the third and final test. After this test was completed, the participants were given the retrospective questionnaires to fill out and return to the researcher.

Several problems related to the timing of the tests arose during the testing period. First, not all participants completed the game within the four allotted weeks and were therefore slightly too late with completing all three tests. Second, one participant neglected to complete the second test immediately after finishing the game, and due to the timing of the experiment was therefore asked to only complete the third test and fill out the retrospective questionnaire.

4.4: Design and Analyses

After scoring and coding the results of the tests outlined above, several different variables related to the constructs discussed previously were investigated using different statistics and methods.
4.4.1: Quantitative and statistical analyses

4.4.1.1: Normality of distribution analysis
Before starting the analysis on the quantitative data, a normality of distribution analysis was performed on the Isdt and chunk completion data, in order to determine whether parametric or non-parametric statistics were to be used. Because several tests for normality of distribution, discussed in more detail in the results section, showed that much of the data is not normally distributed, non-parametric statistics were chosen to investigate the different relations between variables.

4.4.1.2: Y/N test results analysis
Two different analyses were performed on the re-calculated results of the Yes/No tests, one related to the construct of overall increase of receptive vocabulary and one related to the construct of frequency of occurrence of vocabulary. The alpha-level for each analysis performed on the Isdt or chunk completion test data was set at p < .05. For the first analysis, the independent variable is the test moment, which has three levels, namely the pre-test, the post-test, and the delayed post-test, labelled T1, T2, and T3 respectively from this point onward. The dependent variable for each of these tests is the Isdt score obtained by each subject. Because the goal was to see whether there was any significant development between the three tests, a Wilcoxon signed-ranks test is used. This test can be used because the same group of participants was tested at several points in time (Lowie & Seton 2012). Three statistical hypotheses were formulated to test the first main hypothesis posited in chapter 3.4.

- The Isdt scores of the participants for test 2 are significantly higher than those for test 1
- The Isdt scores of the participants for test 3 are significantly lower than those for test 2
- The Isdt scores of the participants for test 3 are significantly higher than those for test 1

For each of these hypotheses, the corresponding null-hypothesis is one of no difference (e.g. ‘The Isdt score of the participants for test 2 is not significantly higher than those for test 1’).

The second analysis performed on the Isdt data resulting from recalculation of the Y/N test data, is a comparison of increase or decrease between different tests per frequency category. For this second analysis the combination of variables is slightly more complex. Unlike the first test, the independent variable is no longer the test moment, but the difference between Isdt scores for test 2 and test 1 per frequency group. This means it has four levels: very high frequency (VHF), high frequency (HF), threshold frequency (TF), and low frequency (LF). Like the previous analysis, the Isdt score is again the dependent variable. Like the previous test, Wilcoxon signed-ranks tests were used to determine whether the change between test 1 and 2 differs significantly between the different independent variable levels. In order to test the full range of distinctions between frequency categories,
the significance of the $I_{sdt}$ score of each frequency category must be tested in relation to those for lower frequency categories. This results in the following statistical hypotheses:

- The difference in $I_{sdt}$ score between T1 and T2 for VHF is significantly greater than HF
- The difference in $I_{sdt}$ score between T1 and T2 for VHF is significantly greater than TF
- The difference in $I_{sdt}$ score between T1 and T2 for VHF is significantly greater than LF
- The difference in $I_{sdt}$ score between T1 and T2 for HF is significantly greater than TF
- The difference in $I_{sdt}$ score between T1 and T2 for HF is significantly greater than LF
- The difference in $I_{sdt}$ score between T1 and T2 for TF is significantly greater than LF

4.4.1.3: *Chunk completion test results analysis*

In addition to the analysis of the Y/N test results, the data resulting from the chunk completion test were analysed in a similar manner. Two different analyses were performed with regard to the constructs of overall increase and frequency category.

For the first analysis, the independent variable is the test moment, with three levels (T1, T2, T3), and the dependent variable is the number of chunks correctly completed. For this test, again a Wilcoxon signed-ranks test is used. The following statistical hypotheses for this test can then be formulated:

- The number of correct completions at T2 is significantly higher than at T1
- The number of correct completions is significantly lower at T3 than at T2
- The number of correct completions is significantly higher at T3 than at T1

For the second within-subject analysis, the independent variable is the difference between test 1 and test 2, with two levels: high frequency and low frequency The dependent variable is again the number of correct chunk completions. For this test, once again a Wilcoxon signed-ranks test was used. The statistical hypothesis corresponding to this test can then be formulated as: the difference between test 1 and test 2 for the number of correct high frequency chunks is greater than that for low frequency chunks

4.4.2: *Qualitative analysis*

In addition to the analyses outlined above, the qualitative data gained from the retrospective questionnaires was analyzed. After the assignments had been coded, the results were tallied and common themes in the participants’ responses were identified.
Chapter 5: Results

For the quantitative data, statistical analyses were performed using SPSS 19. First, some raw data are presented and briefly discussed. Next, the results of the statistics performed on the $I_{sd}$ and chunk completion test data are presented. Finally, the retrospective questionnaire results are presented in the form of a tally of common themes emerging from these qualitative data and a brief discussion of each of these themes.

5.1: Quantitative data: $I_{sd}$ and chunk completion test scores

5.1.1: Descriptive statistics

Before using inferential statistics to assess relationships between different variables, the descriptive statistics of the data are examined to gain a general understanding of the data.

Starting with the $I_{sd}$ data, the table below shows the number of items checked by each subject per frequency range, the totals for each test and the $I_{sd}$ scores for each test.

<table>
<thead>
<tr>
<th>Test</th>
<th>VHF</th>
<th>HF</th>
<th>TF</th>
<th>LF</th>
<th>PW</th>
<th>Total - PW</th>
<th>Total</th>
<th>Total $I_{sd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1</td>
<td>15</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>18</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>18</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>T1</td>
<td>20</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>20</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>20</td>
<td>17</td>
<td>19</td>
<td>7</td>
<td>12</td>
<td>63</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>T1</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>20</td>
<td>14</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>T1</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>19</td>
<td>10</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>13</td>
<td>6</td>
<td>72</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>T1</td>
<td>20</td>
<td>16</td>
<td>19</td>
<td>8</td>
<td>3</td>
<td>63</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>10</td>
<td>6</td>
<td>69</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>T1</td>
<td>20</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>18</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>T1</td>
<td>20</td>
<td>13</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>20</td>
<td>17</td>
<td>18</td>
<td>8</td>
<td>14</td>
<td>63</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>20</td>
<td>14</td>
<td>14</td>
<td>2</td>
<td>7</td>
<td>50</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>T1</td>
<td>16.7</td>
<td>6.1</td>
<td>6.1</td>
<td>1.7</td>
<td>0.7</td>
<td>30.7</td>
<td>31.4</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>18.8</td>
<td>11.2</td>
<td>10.0</td>
<td>3.8</td>
<td>4.3</td>
<td>43.8</td>
<td>48.2</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>19.1</td>
<td>11.3</td>
<td>10.6</td>
<td>4.1</td>
<td>4.3</td>
<td>45.1</td>
<td>49.4</td>
</tr>
</tbody>
</table>
Several interesting issues can be observed in these data. First of all, a first look at the data shows substantial variability between the participants in terms of the increases between tests, the total number of items checked and the number of pseudo-words they checked. This last factor in particular has a significant impact, which is clearly visible in subject 2’s score on the third test, in which she indicates knowing more words than on test 1 and 2, but due to having checked 12 pseudo-words has a lower $I_{nd}$ score connected to it. Secondly, a similar observation can be made about the mean scores at the bottom of the table, we can see that while there appears to be an increase in the number of items checked by the participants between each test, the mean $I_{nd}$ scores for each test do not reflect this pattern.

Next, examining the chunk data also reveals several interesting developments, presented in the table below.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Delayed post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>Total</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

The table and chart show a substantial amount of variation in the data, both within and between participants. The most visible trend is that of a U-shaped development exhibited by the data from
participants 2, 4 and 5, indicating that the number of correct chunk completions dropped after playing the game. Another trend is visible in the data from subject 3 and 7, which shows no difference between the first two tests and only rise at the third test. However, many of the differences consist of one or two points, the statistical significance of which is discussed below.

5.1.2: Inferential statistics

5.1.2.1: Checking assumptions of normality of distribution
The normality of distribution of all quantitative data was checked using Shapiro-Wilk analyses, since this study used a sample size smaller than 50 (Lowie & Seton 2012). For the first test using the Isdt data, a Shapiro-Wilk test showed that the Isdt data for test 2 and 3 were normally distributed at p > .05, while the data for test 1 was not normally distributed. For the second test using the Isdt data of the different frequency levels, the Shapiro-Wilk test showed half of the data sets for each frequency range to be normally distributed at p > .05 and the other half of the data to be abnormally distributed at p > .05. For the first test using the chunk completion test data, only the dataset for test 1 was shown to be abnormally distributed by a Shapiro-Wilk test at p < .05, whereas the other two data sets were shown to be normally distributed at P < .05. Finally, for the second test using the chunk completion test data, only the first dataset (indicating the difference between the scores on high frequency chunks between test 1 and 2) was shown to be normally distributed, while the second dataset was not normally distributed. In light of the different distributions, non-parametric statistics were used for all of the quantitative data presented in this study. Because data from the same group at different times is analyzed, a Wilcoxon signed-ranks test is used.

5.1.2.2: Isdt scores
Starting with the analysis of the overall Isdt scores, a Wilcoxon signed ranks test showed that the overall difference between Isdt scores of the participants at the time of the pre-test, post-test and delayed post-test were not significant at p < .05. Next, the same test was performed on the differences in Isdt scores between the pre-test and post-test, per frequency group. This analysis revealed that only the difference between the Isdt scores for the pre-test and post-test between the high frequency (HF) and the low frequency (LF) test items was significant at p < .05 (Z = -2.201, p = .028). This means that the increase in Isdt score resulting from the participants’ performances on the pre- and post-test was significantly higher for the HF items than for the LF items. All other relationships between the Isdt scores in different frequency ranges were not significant at p < .05.

5.1.2.3: Chunk completion test scores
Beginning with examining the overall increase of the number of chunks the participants completed correctly, a Wilcoxon signed ranks test showed that only the difference in the number of chunks
correctly completed between the pre-test and delayed post-test was statistically significant at $p < .05$ ($Z = -2.06$, $p = .039$). This means that while the number of chunks correctly completed between the pre- and post-test, and between the post- and delayed post-test were not significant at $p < .05$, there is a statistically significant increase between the pre-test and delayed post-test. Additionally, the test showed that the difference in change between the two tests per frequency group (i.e. high frequency and low frequency) was not significant at $p < .05$.

5.1.2.4: Summary

In summary, the application of Wilcoxon signed ranks tests to the Isdt and chunk completion scores for the three different tests, and the differences between the first two tests per frequency group for both Isdt and chunk completion measures showed that the only significant results can be found in the difference between the Isdt scores at test 1 and test 2 of HF and LF test items, and between the chunk completion scores at test 1 and test 3.

5.2: Qualitative data: retrospective questionnaires

Examining the qualitative data gained from the retrospective questionnaires completed by the participants resulted in the identification of several common themes present in the writings by the participants, relating to several of the factors discussed in the background.

Beginning with the issue of context, five of the six participants indicated being able to learn words from the situated context of the game, and two indicated specifically being able to connect words to images. However, four participants indicated that the language in the game was not comprehensible (enough) to them and hence four of them also indicated that they found learning vocabulary by lexical inferencing to be impossible. While this may not have been possible, all but one of the participants did find the use of subtitles and closed captions to be helpful, with participants finding them useful for observing word boundaries and recognizing words by spelling.

In addition to using the context, all six participants indicated being able to make connections between certain single Spanish lexical items and similar words in other languages they knew. Languages mentioned in this regard include French, Italian, Latin, and English.

Four participants indicate being able to relate some of the game’s language to their pre-existing knowledge, only all agree that this is far too limited to make any significant contribution to their understanding of the game. Only one participant, namely number 7, indicated being able to make use of her Spanish pre-knowledge, which was also more extensive than that of the other participants completing the retrospective questionnaire.

Finally, while only two participants specifically refer to extraneous cognitive load by indicating that they found controlling the game and attending to the language used in it to be difficult, five of the six participants indicated initially trying to pay attention to the language in the game, but
gradually finding themselves unable to do so after some time. The reason for these problems with attention is a general demotivation to follow the plot of the game due to the incomprehensibility of the TL and the high speed at which it is spoken by NPCs.
Chapter 6: Discussion

Introduction

In this chapter the results presented in chapter 5 are discussed in the context of the literature presented in chapter 2 and 3, and the hypotheses posited in chapter 3.5 are assessed.

6.1: Incidental FL vocabulary acquisition from COTS video games

Only limited support was found in the results for the first hypothesis examined in this study. Although the results in the previous chapter show that significantly more Spanish FL chunks were correctly completed by participants at the time of the delayed post-test than at the time of the pre-test, the analysis of overall increase between tests of the I_{alt} scores did not show any significant results. The main reason for this limited support for the first hypothesis may be a flawed test design, which is discussed in more detail below.

One possible explanation for the high number of checked pseudo-words and concurrent non-significant I_{alt} scores for the first vocabulary test may be found if a dynamic perspective is taken with regard to the data. As was discussed briefly in chapter 2.5, several DST-based studies have found language learners overgeneralizing newly acquired forms during initial periods of high variability in development (Spoelman & Verspoor 2010; van Dijk, Verspoor & Lowie 2011). Based on the observations made of the high between-subject and within-subject variability in the I_{alt} score results, we might be able to speculate that, at these three initial points of acquisition, the participants may overgeneralize the approximate forms of FL vocabulary they encountered in the game.

6.2: Factors facilitating incidental FL vocabulary acquisition from COTS video games

Supplementing the first hypothesis, the second hypothesis was posited to investigate the possible facilitating effects of frequency of occurrence, context and game enhancement through subtitles.

The statistical analyses of the quantitative data found limited support for hypothesis 2A, which states that more high frequency FL vocabulary is incidentally acquired from COTS video games than low frequency vocabulary. The data partially support this hypothesis, showing that the number of lexical items known to the participants in the high frequency category grows significantly faster than that in the low frequency category. However, the support is partial because no other relations between frequency categories were found to be significant. Again, a flawed test design may have contributed to these results, but they do clearly demonstrate that the question of what kind of impact frequency has on the learning of language from video games remains one which merits further investigation.
The other two sub-hypotheses of hypotheses 2 did find some support in the form of the participants’ reflections on their importance in their retrospective questionnaires. Partial support can be found for hypothesis 2B, which states that the context, in terms of textual, situated and embodied context, in which foreign language vocabulary in a COTS video game occurs, facilitates incidental acquisition of that vocabulary. The fact that all but one of the participants perceived situated knowledge to be important in their acquisition of language from the game underscores the importance of situated and embodied meaning in learning from video games and in learning in general, as is argued by Gee (2007) and Neville (2010), among others. However, most of the participants did find the language in the game too incomprehensible to enable lexical inferencing, which means that this hypothesis is only partly supported by the data. This is despite the fact that the game’s comprehensibility was enhanced using subtitles, which were widely seen as being helpful. The fact that most of the participants found subtitles and captions to be helpful for identifying word boundaries and for learning the form of words, lends support to hypothesis 2C.

6.3: Factors inhibiting incidental FL acquisition from COTS video games

Contrasting with the previous section, support was found in the qualitative data for hypotheses 3A and 3B. Starting with hypothesis 3A, four of six participants indicated not being able to follow the language in the game due to their negligible proficiency in the target language. This supports the hypothesis stating that the low level of FL proficiency of the participants inhibits or prevents the incidental acquisition of vocabulary from a COTS video game in that foreign language. However, there was one subject (number 7) who indicated being able to understand the language in the game enough to enable her to use lexical inferencing. This is likely due to the fact that she had received some prior instruction in Spanish, which seems to support the statement made by Rankin, Gold and Gooch (2006) about games lacking flexibility to provide good learning environments for learners at different proficiency levels.

Connected to this issue of problematic comprehensibility is the issue of extraneous cognitive load and problems with attention. This is where we find the most unanimously agreed-on point made by all of the participants in the data. While all of them tried to pay attention to the language when they started playing the game, their attention soon lapsed due to both the incessant stream of incomprehensible language coming their way, and the extraneous cognitive load placed on them by having to control the game. This means that, at least in the experience of several participants, the extraneous cognitive load prevented them from paying attention to the language in the game. This supports hypothesis 3B and substantiates the earlier research into this problem highlighted in section 3.3.2, specifically deHaan, Reed and Kuwada’s (2010) findings of the detrimental effect of interactivity on noticing and recall of new vocabulary.
6.4: Limitations of the study

This study has several limitations in terms of flaws in the research design and its overall limited scope.

Starting with the research design, certain choices made by the researcher may have affected the participants’ behaviour on the tests, which in turn may have resulted in the fact that almost no significant results were found, especially for the I_{sel} scores. The most important of these flaws was the initial inclusion of a confidence interval for each yes/no answer given by the participants. For each question, they were asked to indicate how sure they were about their answers. The phrasing of these questions resulted in confusion for some participants and was therefore not included in this thesis.

Additionally, the lack of measures to help those participants with little or no gaming experience cope with control of the game can be considered a flaw in the study’s design. While the idea of organizing a pre-playing meeting with the participants to help them install and start playing the game under supervision of the author was considered, it was not implemented in the final research design. This could have helped inexperienced participants cope slightly better with the cognitive load placed on them, by learning to control the video game before the actual playing started. However, this would not have eliminated all sources of cognitive load, only those associated with mastering control over the video game.

Finally, the study’s limited scope in terms of sample size, the aspects of vocabulary under investigation, and the duration of the study may have contributed to the lack of significant results. Although time constraints prevented using a larger sample than the seven participants who took part in the investigation, the results would be more representative with a larger group of participants. Similarly, while these time constraints also limited the amount of time spent on playing video games, the participants only spent a relatively short amount of time playing the video game. Even though some participants reported spending more than 50 hours on finishing the game, what has to be considered is the fact that a gamer who averages 10 hours of gaming per week will average more than 500 hours per year. The gaming statistics presented in the very beginning of this study showed that there are large groups of people (at least in the USA) who spend even more of their time gaming. Consequently, this study was only able to capture a brief glimpse into the foreign language development that could take place during the many hundreds of hours of game playing by so many people around the world.
Chapter 7: Conclusion

This study provided little in the way of empirical support for the hypotheses presented above, but it did manage to provide some understanding of the perspectives of language learners on the ways in which playing a video game may help or hinder their acquisition of the language.

This means that there appears to be a complex answer to the main research question of this study: is incidental acquisition of foreign language vocabulary from video games possible? Although only limited empirical evidence resulted from this study supporting the argument that this is possible, the qualitative data did show that the people partaking in the experiment found it to be possible, albeit marginal.

Despite these answers, investigation into the educational potential of all kinds of video games remains a promising field for future studies and experiments, due to complex and dynamic interaction of different factors, some of which may be unique to video games. In the author’s opinion, future research into this topic may be able to establish empirically and statistically whether vocabulary can be acquired from FL video games, and more specifically what factors contribute to this. The interactivity inherent in video games means that they are unique when compared to other media, and therefore further research into situated and embodied learning from video games could yield some interesting results. While such researchers as Gee (2007) and Neville (2010) have made some initial steps in this direction, a more comprehensive investigation of the role of learner agency and situated knowledge in video games specific to language learning could provide some interesting insights. Additionally, examining more closely certain individual differences between learners, in terms of gender, motivation or language learning aptitude may be able to find those factors which, combined with playing video games, can facilitate foreign language acquisition. Finally, in relation to the limitations of this study discussed above, more comprehensive studies involving a larger and more diverse group of participants, a longitudinal research design and testing for the acquisition of different aspects of a foreign language, and pre-playing instructions for inexperienced players, may yield a more comprehensive and ecologically valid account of the possibilities of COTS video games in foreign language learning.
References


