Enhanced incidental learning of formulaic sequences by Chinese learners of Italian

by

Ilaria Borro

The thesis is submitted in partial fulfilment of the

requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

of the University of Portsmouth

January 2021

Abstract

The present study has two main aims, on both a pedagogic and theoretical level. The pedagogic focus is to provide a theoretical and experimental basis for instructional techniques that trigger enhanced incidental learning. The rationale for enhanced incidental learning lies in the need to improve L2 speakers' implicit knowledge, which requires keeping learning as incidental and unconscious as possible. At the same time, purely incidental conditions result in slow and limited knowledge gains, and this shows the need for devices capable of speeding up acquisition while simultaneously keeping the learner's level of consciousness below the awareness threshold. Enhanced incidental learning is such a device.

The need to verify whether such conditions are capable of triggering genuine incidental learning and resulting in implicit knowledge leads to the second aim of the study. This focuses on a more theoretical and psycholinguistic issue: the relationship between the level of consciousness at the point of learning and the kind of knowledge gained. As a corollary of this investigation, the possibility for learning to take place below the level of awareness is addressed.

The target structure chosen for the experiment is formulaic sequences, since mastering them is considered both a crucial and problematic part of second language acquisition. Notably, the capacity of the human brain to unconsciously tally co-occurrences makes formulaic language an optimal target structure for the investigation of statistical incidental learning.

83 Chinese learners of Italian L2 were exposed to reading-while-listening to a graded reader including seven occurrences of each target idiom. Participants were randomly assigned to one of four experimental groups and exposed to the target items with (i) typographical, (ii) aural, (iii) typographical + aural, or (iv) no additional enhancement. A control group performed the tests with no treatment. Learning was assessed through both offline and online tests, which were performed immediately after the treatment and again three weeks later. To investigate the learners' level of consciousness, a subsample of participants had their eye movements recorded at the process level. In addition, stimulated recalls provided information about participants' awareness of the enhancement devices and the learning task.

Findings show that significant knowledge was gained. Awareness measures show that treatments involving typographical enhancement resulted in intentional rather than incidental learning, which was confirmed by the post-test detecting mainly explicit

knowledge. While confirming the effectiveness of typographical enhancement for explicit knowledge gains, this result shows that learning conditions involving it cannot be considered incidental.

In contrast, subjects exposed to aural enhancement allocated additional attention to the target items without being aware of it, therefore engaging in genuine incidental learning. This process resulted in both explicit knowledge (at the receptive level only) and significantly increased automatic familiarity with the target items. This finding is original and important on two levels. First, it provides evidence for the effectiveness of aural enhancement, which was lacking in the existing literature. Secondly, it supports the possibility for enhanced incidental learning to take place below the level of awareness, thus contributing to a key debate in the SLA field.

Results concerning the increased-frequency-only treatment are not straightforward and therefore need further research.

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word Count: 61696

Contents

List of Tables

List of Figures

List of Abbreviations

AE:	Aural Enhancement
ANOVA:	Analysis of Variance
CSI	Collocational Sensitivity Index
EIL:	Enhanced Incidental Learning
ET:	Eye-Tracking
FFD:	First Fixation Duration
FL:	Formulaic Language
FPFC:	First-Pass Fixation Count
FPRT:	First-Pass Reading Time
FS:	Formulaic Sequences
GSI	Grammatical Sensitivity Index
IE:	Input Enhancement
IFO:	Increased Frequency Only
L1/L2:	First Language / Second Language
NNS:	Non-Native Speaker
NS:	Native Speaker
RT:	Reaction Time
SK:	Skipping rate
SLA:	Second Language Acquisition
SPR:	Self-Paced Reading
TAE:	Typographical and Aural Enhancement
TE:	Typographical Enhancement
TFC:	Total Fixation Count
TRT:	Total Reading Time
WM:	Working Memory

Acknowledgements

Having Professor Mike Long as a supervisor was an incredible honor. His patient and precise guidance throughout the whole process and his passion for research have been and will be a source of profound inspiration. I cannot express how grateful I am for the opportunity of working with him.

I am deeply thankful for Professor Alessandro Benati's supervision: his methodological directions, clear timelines and reassuring attitude have been crucial for navigating through these years.

A mention is due to Professor Stefano Rastelli, for introducing me to Alessandro and Mike.

I also owe to express my appreciation to Dr Nick Bertenshaw, for accepting to be part of my committee even though the project was already defined.

Throughout the completion of this project, I asked for and received help from numerous people, whom I must thank for their kindness.

First, the Marco Polo-Turandot students who accepted to take part to the study as experimental subjects. A crucial help came from my fellow teachers at Università degli Studi di Pavia, who patiently accepted to have their lessons interrupted and used their precious time to answer my questions about their students.

Professor Carla Bagna, Dott. Andrea Scibetta and Dott. Yang Ni from Università per Stranieri di Siena have my heartfelt gratitude for their great help and hospitality when I was collecting data in Siena.

Gaia Chiuchiù deserves special thanks for allowing me to pilot my study in her school and for being of constant support and encouragement throughout the project: one could not ask for a better peer to face this process with.

I owe huge gratitude to Professor Marco Porta, for lending me so graciously the eye tracker and the lab, for patiently training me and for his unvaluable work with the raw data.

Beyond this, I am thankful to the friends/colleagues in the SLA program who inspired me and helped me with encouragement, advice and stress-management strategies.

Lastly, my gratitude goes to my family: to my husband, for his practical support, and to my children, Michelangelo and Maria Caterina, for their patience and forgiveness towards my physical and emotional absences.

Dissemination

Second Language Research Forum 2020, Vanderbilt University

23-25 October 2020

Presentation: Enhanced incidental learning of formulaic sequences from reading. A SPR and ET study.

XX Congresso internazionale dell'Associazione Italiana di Linguistica Applicata, Università per Stranieri di Siena

20-22 February 2020

Presentation: Arricchimento grafico e acustico dell'input e acquisizione incidentale di espressioni formulaiche. Uno studio con apprendenti sinofoni di italiano L2.

EuroSLA 29, University of Lund

28-31 August 2019

Presentation: Enhanced incidental learning of formulaic sequences by Chinese learners of Italian.

LAEL International Postgraduate Conference, University of Lancaster

11 July 2019

Presentation: Unobtrusive input enhancement and incidental learning of multi-word units.

Vocab@Leuven, KU Leuven

1-3 July 2019

Poster: Borro, I., Malone, J. & Long, M. Enhanced incidental learning of vocabulary and formulaic language.

BAAL Annual Conference, University of York

6-8 September 2018

Presentation: Unobtrusive input enhancement and incidental learning of multi-word units. Preliminary results.

BAAL Vocabulary SIG Conference, University College London

9-10 July 2018

Poster: Unobtrusive input enhancement and incidental learning of multi-word units. An exploratory study with Chinese learners of Italian L2.

Workshop: "Language Development in Different Contexts", University of Reading

17 May 2018

Presentation: Unobtrusive input enhancement and incidental learning.

Introduction

Formulaic language constitutes a riddle for second language acquisition researchers and practitioners. A wide range of pedagogic techniques has been employed with little success, with longitudinal studies demonstrating the poor performance in this area even of advanced learners.

While explicit teaching is unviable, due to the huge number of formulae languages comprise, their statistical dispersion and low salience make learning from natural input slow and ineffective. A new perspective on teaching formulaic language is clearly needed, and enhanced incidental learning constitutes a proposal in this respect. By combining increased salience with incidental learning conditions, the pedagogic technique addressed here aims at fostering unconscious learning and the acquisition of implicit knowledge.

The key consideration leading to the formulation of such a goal is the acknowledgement of implicit knowledge as the main objective for language teaching. Even though explicit knowledge is capable of aiding the process of language acquisition, it is widely recognized that the speaker needs implicit knowledge for online communication, i.e. listening, speaking, and other real-time language use.

The aim of the present work is therefore to design and test enhanced-incidental-learning conditions, verifying their effects at two stages. At the process level, the learners' consciousness during the pedagogic treatment is monitored, in order to investigate the intervention's ability to trigger implicit learning processes. At the product level, both implicit and explicit knowledge gains are measured, to verify the effectiveness of the learning conditions. In addition, collecting data at both the process and product level helps address a debated issue, i.e. the possibility for learning to take place below the level of awareness or, in more general terms, the relationship between the level of consciousness at the point of learning and the kind of knowledge gained. Such investigations are only possible through triangulating different awareness measures and employing online assessment tools. Therefore, the present study includes eye-tracking, retrospective verbal reports, and self-paced reading, as well as a measure of explicit learning.

The way the present research aims were driven by theoretical notions and existing research lacunae, the methodological choices, the dependent measure outcomes, analyses, and implications are described and organized in the thesis as follows.

In the first chapter, the priority of implicit over explicit knowledge as a goal for language teaching is justified. Given that instructors should aim for implicit knowledge, the chapter then addresses how to accomplish such a goal. The literature about the interface debate demonstrates that explicit knowledge cannot turn into implicit knowledge, which implies the need for implicit learning to take place directly. Creating conditions for incidental learning is therefore required of the instructor wishing to trigger implicit knowledge gains.

The second chapter deals with the choice of formulaic language as a target for the present investigation, by reporting evidence of its importance and difficulty. Moreover, theoretical and empirical literature about the effects of incidental learning conditions on L2 formulaic sequence learning is accounted for, demonstrating why it is reasonable to expect enhanced incidental learning to be effective with this kind of structure.

The third chapter focuses on enhanced incidental learning. The first section describes its theoretical and psycholinguistic underpinnings. Then, the pedagogic tools chosen in order to create enhanced-incidental-learning conditions are described. The empirical literature demonstrating their potential effectiveness for the study's aims is reported, as well as the gaps and methodological flaws that the present study seeks to overcome.

Chapter four reports the methodology adopted in detail, including the composition of the sample, the selection of target items, the instructional material, the testing instruments, and the procedures followed.

The fifth chapter reports the statistical analysis of the data, which are organized in two different experiments. The analysis in Experiment 1 deals with the whole sample, the post-tests and the retrospective verbal reports. Experiment 2 focuses on the subsample which performed the pedagogic treatment while having eye movements recorded by means of eye-tracking. Therefore, in addition to the post-tests and outcomes of the retrospective interviews, for this part of the sample the eye-tracking measures are analyzed, as well.

The sixth and last chapter reports the result discussion and implications, as well as limitations and the directions for subsequent research.

Chapter One: Theoretical Background

The chapter aims to define the theoretical framework for the present study and to show how this empirical research relates to it. To achieve this goal, some core theoretical notions are described and connected.

The first assumption in need of justification is that acquiring implicit knowledge should be the main goal of language pedagogy. In order to affirm this, the most relevant definitions of implicit and explicit knowledge are reported and the debate about a possible interface between the two is described (section 1.1).

The core theoretical notion necessary to understand the effectiveness of pedagogic interventions aimed at creating implicit knowledge is that different levels of consciousness can be activated during the learning process. The second section (§ 1.2) addresses these levels and the notion of salience, i.e. the main factor capable of affecting consciousness. In relation to levels of consciousness and salience, the key concepts of implicit and explicit learning are defined.

Despite the differences among the models proposed in the literature, it can be affirmed that the term 'implicit learning' refers to an internal, unconscious process, which cannot be totally under the instructor's control. In order to investigate implicit knowledge, what an external intervention can aim for is just setting the optimal conditions, which are known as incidental-learning conditions. Therefore, the third section (§ 1.3) addresses the notion of incidental learning, and briefly introduces some methodological issues driving the design of the present study.

1.1. Implicit knowledge is the goal

1.1.1. The nature of implicit and explicit knowledge

There is broad agreement in the SLA literature about the existence of two different types of linguistic knowledge, usually known as explicit/learned and implicit/acquired knowledge. As Whong and colleagues (2014) observe, this distinction is a point of agreement even between the two main approaches to contemporary research on SLA: the generativist approach and the cognitivist approach. Indeed, second language researchers agree about the difference between a conscious and controlled type of knowledge and a subconscious and automatic type of knowledge. However, in the last decades numerous definitions have appeared in the literature, often diverging in some respects, especially with regard to the possibility of one kind of knowledge being transformed into the other. These points (i.e. the different natures of learned and acquired knowledge and the relationship between them) are especially relevant pedagogically, because they can be considered a key to defining the goals and interpreting the effects of language teaching.

The definition of two different kinds of knowledge in SLA dates to the late 1970s. Krashen (1977, 1982) distinguishes between learned and acquired knowledge. According to his definition, acquired knowledge (also called implicit knowledge and acquired competence) is subconscious, and it is created and used without awareness, as children do with their L1; it is related to fluency and to the intuitive feeling of correctness all native speakers have about their first language. Bialystok (1978) was one of the first to adopt Krashen's definitions. She states that implicit knowledge is intuitive, automatic, spontaneous and used to produce or understand the target language; as such, it is related to fluency and to immediate and spontaneous language responses (production or comprehension). According to Reber (1989), tacit (i.e. implicit) knowledge has three defining qualities: it is reasonably veridical, as long as it mirrors the invariances of the input; it is partial because it does not include all the regularities in the environment; and it is structural, in that the patterns are manifestations of the abstract rules underlying the input.

On the other hand, learned knowledge (also called explicit or formal knowledge) is the result of a conscious process of learning about the language. It includes metalinguistic notions commonly known as 'grammar' or 'rules' (Krashen 1982). It contains what the learner consciously knows and can report about the language; it comprises grammar rules and formal aspects of the language, but it does not imply the ability to use the language effectively. Explicit knowledge is related to language responses that are deliberate and (even very briefly) delayed (Bialystok 1982).

The existence of a difference between implicit and explicit language knowledge is confirmed by brain science, which, through investigations of brain damage, demonstrates that implicit and explicit memory have anatomically distinct neural substrates, i.e. they are stored in different areas of the brain (A. W. Ellis & Young 1988). A key contribution to the issue is that of Paradis (1994, 2004), who provides greater depth in discussing the nature of the two kinds of knowledge:

> "Explicit knowledge is qualitatively different from implicit competence. Explicit knowledge is conscious awareness of some data (utterances) and/or of their explicit analysis (structure). Implicit competence, on the other hand, is a set of computational procedures (of which the speaker is unaware) that generates sentences (which serve as data from which linguists or reflective speakers may construct a grammar—a set of rules—that becomes part of one's explicit knowledge)."
>
> (Paradis 2004, p. 47)

In other words, the very objects of implicit and explicit knowledge are different: explicit grammar rules describe the language, but they do not correspond to the implicit representations the brain relies on in order to produce utterances. The computational procedures that generate sentences are inaccessible to conscious observation, cannot be explained, taught, or intentionally learned. Consequentially, metalinguistic notions cannot be based on the implicit linguistic competence, which always remains covert; instead, they are abstractions derived from the linguistic output created according to the (unobservable) implicit representations.

Similar ideas are maintained in the generative framework. Schwartz (1993) employs the terms competence and learned linguistic knowledge when making the distinction between mental representations (i.e. what has been referred to as implicit knowledge) on the one hand and knowledge based on textbook rules (i.e. explicit knowledge) on the other. As in the frameworks exposed above, competence cannot be learned explicitly, but only through the processing of input. Specific to the generative framework, such processing needs to be coupled with Universal Grammar, a language-specific internal mechanism.

To wrap up, the distinction between implicit and explicit knowledge has been widely (though not universally, see e.g. DeKeyser 1994) adopted and acknowledged since the 1980s, and clearly imply a fundamental corollary: the main goal of language teaching should be the learner's acquisition of implicit knowledge (Long 2017). This statement, as the difference existing between implicit and explicit knowledge, finds wide consensus in the literature, bringing together researchers from both generativist and general cognitive paradigms (Whong et al. 2014, VanPatten, Smith & Benati 2020). Indeed, desirable aspects of language use, such as fluency, spontaneous speech and listening comprehension, operate as a function of implicit competence (Bialystok 1978; Long 2017; Whong et al. 2014). Moreover, implicit knowledge is deeper, more durable, faster and more efficient. Being automatic, it is less vulnerable to the effects of a concomitant memory load, fatigue, stress or noise (Paradis 2004) or, as Reber (1989, p. 14) puts it, "implicit systems are robust in the face of disorders that are known to produce serious deficits in conscious, overt processes." Thus, when a learner uses implicit knowledge of language, more attentional resources are available for concentrating on message content and on the actual goal of communication.

1.1.2. The interface debate

Although the priority of implicit over explicit knowledge is widely acknowledged, there exists less agreement about how to acquire it. As Sharwood-Smith stated in 1981: "The ultimate, most highly prized goal of learning, i.e. spontaneous, unreflecting language use, is uncontroversial. How this is achieved is, of course, a matter of considerable debate" (p. 159).

This debate involves numerous issues, one of the most central being the existence of an interface between explicit and implicit knowledge, i.e. the possibility of explicit knowledge becoming implicit. This point is fundamental in regard to a core pedagogic controversy: the effectiveness of formal classroom instruction. Essentially, if explicit knowledge of rules cannot turn into the desired implicit competence, then traditional grammar instruction is probably not an effective use of limited classroom time. On the other hand, if explicit knowledge can either become, or aid the development of, implicit knowledge, then explicit teaching of grammar is validated and justified.

Several models of language learning account for such an interface. In Bialystok's (1978) model, for instance, implicit knowledge is a "working system" containing the information necessary for spontaneous language use. On the other hand, explicit knowledge recognizes and stores all the new information the learner meets; some of this information always remains represented as explicit, whereas part of it can "after continuous use, […] become automatic and transferred to Implicit Linguistic Knowledge" (p. 72). At the same time, implicit knowledge can become explicit, through conscious analysis and observation of one's own linguistic output. Bialystok thus holds what has been called a 'strong-interface position', advocating the possibility of explicit knowledge becoming implicit through practice.

Many scholars in the following years took similar positions (e.g. Anderson 1982; McLaughlin et al 1983). DeKeyser in 1994, advocating to what is referred to as Skill Acquisition Theory (SAT), claimed that explicit knowledge can become proceduralized and then automatized through practice. According to this view, language acquisition happens through the same process as any other human skill, from computer programming to riding a bike. It first requires consciously controlled performance, which relies on explicit memory; the controlled performance is then automatized through practice, thus becoming implicit. If this is the case, explicit grammar instruction should, through practice, result in implicit language competence. This would furthermore justify traditional formal instruction, because the explicit knowledge it creates becomes implicit. Although admitting in later works (2003; 2015) that explicit knowledge may not be directly transformed into implicit knowledge, DeKeyser maintained the idea that the former is necessary for the creation of the latter.

A very different claim is the 'non-interface position', advocated by scholars like Krashen (1982) and Paradis (1994; 2004). In Krashen's system, clear roles are assigned to the two types of knowledge: acquired knowledge is related to fluency and starts the utterances produced by the speaker, while the only role for learned knowledge is that of a 'Monitor' (the Monitor Hypothesis). The Monitor checks utterances after they are produced by implicit knowledge and makes changes to the output according to stored grammar rules. However, it is possible for the Monitor to be effective only when three conditions are met, e.g. during an unspeeded, discrete-point grammar test: (i) the language performer has time to use the Monitor (a normal conversation is too fast for the Monitor to work effectively); (ii) the language performer is focused on form; and (iii) the language performer knows the grammar rule involved. Under these conditions, performers can use learned knowledge to produce structures that are not acquired yet. However, Krashen states that the Monitor performs this process only when rules are syntactically and semantically simple; otherwise, only explicit grammar tests can elicit the use of the Monitor (Keyfetz 1978, Houck et al 1978, Krashen et al 1978).

In this model, explicit knowledge cannot become implicit, learned rules are not internalized, and even if explicit knowledge precedes acquisition, it does not determine acquisition. Empirical observations support this claim, showing that acquisition can occur without previous learning: fluent speakers can have very limited knowledge of formal rules (Cohen & Robbins 1976; Stafford & Covitt 1978; Schumann 1978). Moreover, it is very common to find explicit knowledge that never becomes implicit (Krashen & Pon 1975). As a consequence, Krashen states that implicit and explicit knowledge are created through different processes, and that learning has a very limited and indirect role, if any, in acquisition. In Krashen's view, the acquisition of implicit knowledge can take place only when the acquirer focuses on the meaning of comprehensible input (the Input Hypothesis): if the language contains structures slightly beyond the acquirer's level, then understanding the meaning might also bring about a subconscious acquisition of the linguistic form.

Along the same lines is Paradis' (1994; 2004) position. The statement that "metalinguistic knowledge does not become implicit linguistic competence" is a consequence of the assumption that the rules stored in explicit memory are qualitatively different from the implicit computational procedures driving spontaneous language production. Implicit competence is not consciously observable; therefore, it cannot be noticed and learned. It can be developed only through practice, which makes linguistic tasks or task components automatic and thus implicit. Before this happens, the speaker may rely on explicit knowledge. Practice can make controlled processes increasingly fast, to the point that sped-up explicit knowledge can resemble an automatic procedure. Strong-interface advocates claim that this kind of output results from implicit knowledge; however, neurological studies have demonstrated the qualitative difference between the two kinds of knowledge. These studies suggest that explicit knowledge never becomes implicit; rather, implicit procedures are developed separately by practice and eventually replace controlled processes, because the brain uses automatic (i.e. implicit) systems, when available, by default. In this model, although explicit knowledge can indeed facilitate the acquisition of implicit knowledge, it can never become implicit. Rather, the practice of a form through explicit knowledge provides input for the underlying implicit processes to establish themselves.

The generative framework shares the non-interface position (VanPatten, Smith & Benati 2020; VanPatten 2016; VanPatten & Rothman 2014). According to generativists, the actual mental representations constituting language are abstract, complex and very far from the surface descriptions that grammar textbooks provide of them. Therefore, it is not possible for explicit instruction to affect in any way the development of implicit knowledge, since "implicit and explicit knowledge are qualitatively different, […] in terms of their fundamental content." (VanPatten 2016, p.7). Such position is empirically supported by a self-paced reading study investigating the acquisition of L2 Spanish grammar (VanPatten, Keating & Leeser 2012). The online measurement revealed that the intermediate learners were more likely to behaved like native speakers on the structure for which they did not receive instruction. On the other hand, they showed no implicit sensitivity to violations when it came to the structures they were explicitly instructed about.

Besides the strong-interface and the non-interface position, a third paradigm has been proposed in the SLA literature. The 'weak-interface position' acknowledges some of the theoretical assumptions of the non-interface position, yet assigns an important role to explicit instruction. According to this view, explicit and implicit knowledge involve different kinds of representations, stored in different parts of the brain and are therefore qualitatively different; as a consequence, explicit knowledge cannot be converted into implicit. However, in this model, the two kinds of knowledge do interact. While comprehensible input and communicative practice are still of prime importance, explicit knowledge can help considerably, under certain conditions, to promote implicit competence.

Decades of empirical studies, reviews and meta-analyses (e.g. Norris & Ortega 2000; Goo et al 2009), show that explicit instruction can indeed speed up acquisition. Scholars have defined this core concept in various ways (e.g. Bley-Vroman 1988; Hulstij & De Graaff 1994; Long 1991, Schmidt 1990; Sharwood-Smith 1993; Van Patten & Cadierno 1993). Nick Ellis' (1994; 2005) theory is one of the most representative of this position. He claims that most language knowledge is implicit, created through usage and statistical learning:

> "Frequency of usage determines availability of representation. […] This process tallies the likelihoods of occurrence of constructions and the relative probabilities of their mappings between aspects of form and interpretations, with generalizations arising from conspiracies of memorized utterances collaborating in productive schematic linguistic constructions".
>
> (Ellis 2005, p. 306-307)

Nevertheless, this process is not sufficient to master a second language, because many aspects of language are not salient enough to be learned statistically. For these aspects, a first stage of explicit instruction is necessary. The resulting explicit knowledge does not turn into implicit, but rather facilitates the statistical creation of implicit competence through frequency of production and reception. Similar to Paradis, Ellis points out that explicit knowledge is not transformed into implicit even when it becomes automatic: it remains explicit, only resembling implicit competence by means of speed. Only unconscious, statistical learning creates implicit knowledge. The pedagogic intervention can speed up this process through explicit instruction, exaggerated input and corrective feedback, which improve and direct statistical learning.

In sum, the differences between implicit and explicit knowledge and the priority for language pedagogy to attain the former are widely acknowledged. The qualitative difference between implicit competence and explicit notions, demonstrated by both linguistic and neurological studies (Paradis 2004), excludes the possibility of explicit knowledge turning directly into implicit. This undermines claims of the effectiveness of explicit language teaching for language acquisition. As shown, there is broad agreement that involving learners in meaningful use of the language is either the only (Krashen 1982) or the main (e.g. Bialystok 1978; Ellis 2005; Paradis 2004) way of promoting implicit knowledge, because this triggers the unconscious processes that create representations in implicit memory. Therefore, it can be claimed not only that explicit instruction is not the most effective way of using the learners' time, but also that, over a certain proportion, it can even be detrimental to language acquisition (Long 2017). Focusing the learners' attention on form with little or no communicative value may hamper the unconscious processes that create implicit knowledge when the learners are engaged in meaning-focused activities. On the other hand, it has been observed that a certain amount of explicit instruction can have a positive and relevant role as well, primarily by speeding up acquisition. A deeper insight into implicit and explicit learning, their roles, interaction and the underpinned theoretical notions are the subject of the next section.

1.2. Consciousness: implicit and explicit learning

In the previous paragraph, the priority of implicit over explicit knowledge as a goal for language teaching has been claimed. Moreover, it has been pointed out that in order for learners to gain implicit knowledge, implicit learning needs to take place. In order to provide a more precise definition of implicit and explicit learning two key notions have to be introduced: first, the different levels of consciousness involved in the learning process; second, salience as a chief factor affecting the speaker's consciousness.

1.2.1. The noticing-detection debate

In order to design pedagogic techniques capable of promoting implicit knowledge, it is essential to understand and account for the distinct roles played by implicit and explicit learning. The key notions in this regard relate to the different degrees of awareness and attention involved in learning, i.e. the minimum level of consciousness capable of resulting in new knowledge. The main scholars working on this issue adopted a model proposed in the psychological literature, which distinguishes among consciousness as intention, consciousness as attention and consciousness as awareness (e.g. Schmidt 2010; Tomlin & Villa 1994).

Consciousness as intention addresses the difference between learning a notion deliberately (intentional learning) and learning it while focusing on something else (incidental learning), e.g. learning a grammar structure while reading a text for meaning. The intentional/incidental learning issue is essential to this study and still debated in the literature, as will be addressed in section 1.3 below.

Posner and Petersen's (1990) neurocognitive model describes three different aspect of attention: alertness, orientation and detection, which can be anatomically mapped to different areas of the brain.

Empirical support to this fine-grained analysis of attention is also available in SLA literature, with Leow's (1998) study addressing the role of the three attentional functions in L2 development.

Alertness comprises a "vigilance system", a general readiness to deal with incoming stimuli. In the field of SLA alertness is only of general import and may be related to motivation. Orientation directs attentional resources towards some stimuli and information, excluding others. In SLA, orienting attention towards a part of the input facilitates further processing. Stimuli that were excluded at the level of orientation can still be further processed, albeit with greater efforts.

Detection is the following level of processing, requiring more attentional resources than alertness or orientation. It has been defined as the cognitive registration of stimuli: "Detection is the process that selects, or engages, a particular and specific bit of information. […] Once information is detected, then further processing is possible" (Tomlin & Villa 1994, p. 192). Detection is an essential point of debate about learning and consciousness, because it is the stage at which, some scholars claim, the first opportunity for language acquisition occurs. Namely, according to Tomlin and Villa (1994) tokens are registered in memory through detection and are thus available for learning. In this model, the detection of the functional relationship between linguistic elements is sufficient for the mapping to be learned, and therefore acquisition operates at this level. Crucially, acquisition is assumed to take place without awareness; indeed, none of the three aspects of attention (alertness, orientation, detection) requires awareness to operate. Awareness can be added to them, thus augmenting alertness and orientation, which may in their turn enhance detection and therefore learning. However, in this model awareness is not believed to be necessary for language acquisition. This is a fundamental point with deep pedagogic implications (see infra).

Williams (2005) offers experimental support to Tomlin and Villa's claim. He carried out two experiments based on miniature noun class systems. Subjects were exposed to four artificial one-syllable determiners and were told that the difference between the determiners was related to the distance between the object and the subject of the sentence (two determiners for 'far' and two for 'near'). The subjects were not told that determiners also depended on the animacy of the noun they referred to, and actually corresponded to 'near and living', 'near and not living', 'far and living', and 'far and not living'. The participants were trained on the meaning and use of the artificial determiners; then they performed a task where they were asked to focus on the meaning of the novel words in the context of sentences and indicated whether they meant 'near' or 'far', repeating the sentences aloud and creating a mental image of the situation described. They were exposed to six blocks of 24 of these trials and they were told that a memory test about some of the sentences would be carried out afterwards. However, the actual testing phase (before the memory test) consisted of an exercise where the subjects had to choose the correct determiner for a given noun in a sentence, according to what they felt more appropriate. Each sentence clearly defined whether the object was near or far from the subject and offered the two corresponding determiners as options. Therefore, the near-far criterion could not drive participants' choice. Interviews carried out after the testing phase showed that a large majority of the participants were not consciously aware of the animacy criterion; nevertheless, subjects scored significantly above-chance when choosing the correct determiner for living and not-living objects. A possible interpretation of the results is that the animacy feature was detected in the input, associated with the corresponding determiner and thus influenced the subsequent generalization, all without awareness. Therefore, Williams' studies may support the hypothesis that learning without awareness is indeed possible, i.e. that detection is sufficient for language acquisition, as Tomlin and Villa claim. Findings from this seminal study were confirmed in the following literature, which aimed at replicating its results through refined and expanded experimental designs (Leung & Williams 2011, 2012; Rebuschat & Williams 2012).

This possibility is a primary question in this field of research, as well as for this dissertation.

Many scholars oppose Tomlin and Villa's argument, claiming that awareness is necessary for learning and, thus, that the first possibility for learning is at the next level in the unconscious-conscious continuum, i.e. noticing. This claim is the pivotal idea of the Noticing Hypothesis (Schmidt & Frota 1986; Schmidt 1990, 1993, 2001, 2010). Driven by empirical observations of two case studies, Schmidt claims that "input does not become intake for learning unless it is noticed, that is, consciously registered" (Schmidt 2010, p. 722).

The first case study Schmidt based his claims on follows the evolution of an L1 Japanese naturalistic learner of English. Despite the quick development of communicative and pragmatic competence, the leaner's morphology and syntax remained limited. Schmidt explains that this resulted from a lack of attention to, and reflection on, language form, and an over-reliance on implicit learning strategies, i.e. learning through interaction. The second case is Schmidt's own acquisition of Portuguese. He compared the development of his language performance and the notes he kept in a journal, finding that the forms he did not consciously notice (i.e. not reported in the journal) were not acquired, notwithstanding their frequency in the input, whereas the imperfect suffix –ia, pervasive but unnoticed in the input, was not learned until his Portuguese teacher made him aware of it. Starting from these observations, Schmidt rejects the possibility of learning at the level of detection, and states that input can be turned into intake only in the presence of awareness.

As stated above, the model of consciousness in learning describes it as having three stages: intention, attention (in turn having three aspects: alertness, orientation and detection) and awareness. Schmidt adopts a further distinction between three different degrees of awareness: perception, noticing and understanding. The first level, perception, is considered subliminal – i.e. can take place without awareness – and overlaps with detection. The second level, noticing, is defined as "the conscious registration of attended specific instances of language" (Schmidt 2010, p. 725). Noticing was at first considered necessary and sufficient for learning (Schmidt 1990) while later, in a "weak" version of the Noticing Hypothesis, it was just viewed as facilitative (Schmidt 2010). Indeed, in this model, learning is actually a side effect of attended processing and intake is re-defined as the part of input the learner notices. The third level, understanding, is a higher degree of awareness that includes metalinguistic reflection and conscious generalization across instances, i.e. the recognition of a rule or pattern. Understanding is considered facilitative but not indispensable for learning.

Other scholars have put forth related positions. Gass (1988) proposed a model identifying five processing stages in the conversion from input to output. In this framework, the first stage that allows further processing is called apperception, and it implies noticing (Robinson et al 2012). Robinson (1995) too proposes a hypothesis that can be considered complementary to Schmidt's. In Robinson's model, detection is viewed as subliminal exposure and, as such, cannot have relevant effects on learning, except for few milliseconds. Conversely, detection plus rehearsal – i.e. noticing – is necessary and sufficient for encoding linguistic elements in memory.

The importance of awareness for learning is also acknowledged by Sharwood-Smith, who argues for the effectiveness of pedagogic techniques capable of increasing the level of consciousness (Sharwood-Smith 1981, 1991, 1993; Truscott & Sharwood-Smith 2011; this dissertation, § 1.2.2. and § 3.2). Truscott and Sharwood-Smith (2011) propose a cognitive framework that specifies some of the main ideas connected to consciousness in learning. First, they point out that the literature is vague with regard to the object of awareness: terms such as 'linguistic item', 'input', 'information', and so on are used interchangeably without clear definition. Therefore, they call for a more scientific definition, and argue that what is actually available for consciousness are the mental representations of input and language. Furthermore, they put forward the Activation Hypothesis, which considers activation level the key for awareness: "A representation is conscious if and only if its current activation level is above a given threshold value" (Truscott & Sharwood-Smith 2011, p. 513). On this basis, they propose four levels of processing-awareness. The first is 'subliminal perception', which does not reach awareness and overlaps with Tomlin & Villa's detection; this level is not considered sufficient for learning, because the level of activation is too weak. The second level is 'awareness of input' and corresponds to Schmidt's noticing. This activation level is high enough for the representation to be conscious: however, learning at this stage is unlikely. Truscott and Sharwood-Smith revisit the Noticing Hypothesis, arguing that according to Schmidt's statements, what is consciously attended is only a collection of instances unrelated to any specific form; indeed, Schmidt cannot state that a successful learner is aware of all aspects of a linguistic form. Therefore, he allows that the generalization that allows a learner to incorporate these instances into a grammar takes place by unconscious processes. This being the case, Schmidt's Noticing Hypothesis would not be different from claims about the possibility for learning to take place without awareness. Therefore, in Truscott and Sharwood-Smith's framework, the first level of awareness that allows learning is 'noticing-understanding', which implies that "a representation is constructed as the result of processing that treats it as an instance of a particular form, and it reaches an activation level sufficient for awareness" (Truscott & Sharwood-Smith 2011, p. 520). In other words, existing representations must be active and consciously connected to the perceived representation in order for input to become intake. If additional processing produces more conscious representations of the meaning or form of the noticed representation, then the highest level of processing-awareness is reached: 'conscious understanding beyond the noticed representation'. This last stage is not necessary for language acquisition and is more connected to metalinguistic knowledge.

A different angle on noticing is provided by VanPatten (2015). In his presentation of the foundations of processing instruction, he focused the difference between noticing and processing, maintaining that the former is not necessary for the latter to take place, since being aware of each linguistic information would make processing slow and cumbersome. Moreover, he pointed out that the literature on noticing does not take into account whether the learners create form-meaning links when consciously registering a new item in the linguistic input.

The Noticing Hypothesis and its variants argue against Tomlin and Villa, since they all claim that the detection level of attention is not sufficient, and that awareness is necessary for learning. This position finds empirical support in numerous studies. Among them is Hama and Leow's (2010) extension of Willams's (2005) study, which is described above. Hama and Leow improved on some methodological aspects that, they claim, could have hampered the validity of the original study's results. They worked with the same four-determiner system and the same design, including pre-training instruction (which explained to the subjects the distance criterion and not the animacy criterion), a training task and a testing phase. The main methodological change addressed the measurement of awareness. Williams only used offline retrievals in the form of post-exposure questionnaires to measure awareness. This kind of offline elicitation procedure is considered inadequate for measuring awareness in some of the psychological literature, because memory decay may affect post-exposure verbal reports. Therefore, Hama and Leow added verbal reports from a think-aloud protocol to the design, in order to gather concurrent data at the stage of encoding and during the testing phase. The results show that learners who were unaware at the stage of encoding did not demonstrate any animacy bias in the testing phase when combining nouns with the artificial determiners. This evidence runs counter to the existence of learning without awareness – i.e. at the level of detection (Tomlin & Villa 1994) – while offering empirical support to Schmidt's Noticing Hypothesis (for similar findings, see e.g. Faretta-Stutenberg & Morgan-Short 2011; Leow 2000). However, the online think-aloud protocols share with the post-execution recalls a fundamental limitation: both are based on verbalization in order to distinguish implicit and explicit processes. This has been demonstrated to be an inefficient means of measurement, since "awareness is fleeting and cannot be completely recorded" (Schmidt 1995, p.28, quoted in Rebuschat et al 2015). Moreover, think-aloud protocols have often been criticized because they can interfere with the subjects' mental processes while performing tasks and tests. Empirical support for this claim comes from Rebuschat and colleagues (2015). In their study, three experimental groups were tested on an artificial determiner system: two of the groups thought aloud while the third remained silent. Although all groups showed learning effects, only the silent group was able to generalize the acquired knowledge to novel instances.

The same research team had also demonstrated verbal reports not to be a sufficient tool for awareness assessment (Rebuschat et al 2013). Keeping into account the contradictory findings by Hama and Leow (2010) on the one hand and Williams (2005) on the other, their study aimed at confirming the possibility for adult learners to establish new form-meaning connection in incidental learning conditions. Considering the evidence from verbal reports, the knowledge gains their participants showed appeared to be only explicit in nature. However, this study included an additional measure for awareness, i.e. confidence ratings, which showed that the knowledge gained in incidental learning conditions was at least in part implicit.

Clearly, the methodological issue of effectively measuring awareness during learning processes is crucial to research on noticing and detection, and it needs more sophisticated instruments than verbalization. The present dissertation aims to contribute to this area of research (see § 4.5).

1.2.2. Salience

Irrespective of whether or not awareness is necessary for learning, a key notion in the discussion is salience. Salience is one of the main factors capable of affecting learners' level of consciousness; therefore, it needs to be carefully considered when the aim is the creation of implicit knowledge. Indeed, salience is manipulated as an independent variable in the present experiment, and it therefore needs a thorough discussion touching two main points: first, what is salient, and second, how salience affects language learning.

Salience lacks a clear definition in second language acquisition (Gass, Spinner & Behney 2018), but when descriptions are attempted, they mainly deal with the perception of stimuli. Cho and Reinders (2013) report that "salience refers to the ease with which learners can perceive given input" (p.134). Similarly, Gass and colleagues (2018) assume salience to be "a factor that makes something easier to perceive" (p. 1). Likewise, Wulff (2019) defines salience as the "general perceived strength of a stimulus" (p. 24). In relation to salience, Loewen and Reinder (2011) mention noticeability and explicitness of linguistic input, and in a similar fashion VanPatten and Benati (2010) point to the "degree to which something catches a person's attention" (p. 143).

A number of classifications have been put forward in order to better define salience features and interactions with language perception and learning. A first distinction is that between perceptual and constructed salience. Perceptual salience is intrinsic in the features of a given linguistic form, such as stress, accent, or any perceptual prominence capable of attracting attention. For instance, lexical cues are more salient to language learners than verbal inflections and grammatical markers when it comes to getting semantic information (Benati 2013; Cintron-Valentin & Ellis 2015; VanPatten 2006). Ellis (2018) names this phenomenon psychophysical salience, pointing out that it "arises in sensory data from contrasts between items and their context" (p. 21). This is an aspect of bottom-up processing, because it is the linguistic stimulus that attracts attention.

Salience can also be generated by top-down processing. In this case, the speaker's memory and expectations cognitively pre-activate a stimulus, driving attention towards it. An example of perceptual, top-down salience is grounded salience, which describes the improved noticeability of anything deviating from what is expected or typical. Grounded salience relies on expectations about the language, which are not confirmed in the input and therefore generate surprise. Such expectations can have different sources. In the first stages of language learning, the L1 can generate predictions which are not confirmed in the L2. On the other hand, more advanced speakers can also anticipate language on the basis of the statistical learning developed over their L2 experience (Ellis 2018; Gass 1988).

Perceptual salience can also be affected top-down by sociolinguistic factors, especially when it comes to the meaning of lexical units. Kecskes (2006) extends Giora's (2003) Graded Salience Hypothesis from the L1 to the L2, pointing out that in a second language, as well, "salient meanings are privileged meanings stored in the mind of individuals at a given time in a given speech community." (p. 220). In other words, the most salient meaning is the one the speaker considers the most frequent and probable, regardless of contextual bias or other factors such as non-compositionality. This implies that salience is dynamic, affected by experience and sociocultural factors. Similarly, Ellis (2018) stresses the subjective nature of perception, which can be affected by emotional and motivational factors, thus making a stimulus more salient.

The second category of salience is constructed salience, which takes place when a linguistic stimulus is intentionally made more prominent and noticeable by an external intervention. In instructed second language acquisition, this practice is frequent and motivated by the idea that constructed salience can facilitate noticing of a given item or feature and, therefore, boost the likelihood of further processing and, eventually, learning. Sharwood-Smith (1981) first focused on this concept when encouraging practitioners to create input salience by means of 'consciousness-raising' devices. In classrooms and intervention studies, constructed salience is created through a variety of strategies, ranging from interaction to textual enhancement, from glossing to artificially increased frequency. The present study focuses on frequency and visual and aural enhancement, which are thoroughly dealt with in the chapter covering the theoretical and empirical aspects of input enhancement (§ 3.2). Here, it is necessary to examine in general terms the way salience affects language learning.

Rescorla and Wagner (1972) presented a synthetic and exhaustive expression of the way physical salience, psychological salience and expectations affect learning through an equation which Ellis (2018) considers "the most influential formula in the history of learning theory" (p.23). They related the salience of the cue (a) and the psychological importance of the outcome (b) to the amount of processing required by a stimulus (L) and the associative strength between the cue and the outcome (V). The variation in this associative strength is considered an expression of learning, and it is referred to as dV. All of these variables are connected as follows:

$$dV = ab(L - V)$$

In other words, the amount of learning is proportional to the salience of the cue (a) and to the psychological importance of the outcome (b). Moreover, (L – V) expresses that the more a cue is already associated to an outcome, the less association strength can be created. Conversely, novel stimuli imply a close-to-zero V, i.e. more surprise and therefore more learning (Ellis 2018).

Usage-based approaches to second language acquisition share a similar view, as they regard salience as a key factor for learning. Ellis (2018) maintains the learnability of a construction to be related to three elements: (i) psychophysical salience, (ii) contingency of form-function association, and (iii) learned attention.

Psychophysical salience corresponds to the aforementioned bottom-up aspect of perceptual salience and can be described as the prominence of a linguistic form, which makes it capable of attracting attention. Numerous experimental studies have investigated its relationship with learning. Goldschneider and DeKeyser (2001) carried out a meta-analysis meant to evaluate the degree to which five determinants affected acquisition order of morphemes. According to their results, perceptual salience was the determinant with the highest predictive power ($r = 0.63$), followed by frequency ($r = 0.44$) and morphological regularity ($r = 0.41$). In an eye-tracking study, Cintron-Valentin and Ellis (2015) found the greater salience of lexical over morphological cues to be a significant variable in L2 tense acquisition.

Such considerations are not limited to grammar and morphology. Wulff (2019) considers salience to be directly linked to learnability of unusual formulaic sequences. Along the same lines, Martinez and Murphy (2011) point out that since idioms are composed of known words, their actual meaning easily goes unnoticed. Boers and Lindstromberg (2009) make a similar point and consider this one of the factors making L2 formulaicity hard to acquire.

These findings support the Law of Contiguity (see § 1.2.3, 3.1, 3.2), which claims that patterns need to be registered first (with perceptual salience playing a facilitative role), and then statistical learning based on frequency of usage can take place (Ellis 2001, 2002). In this second phase, contingency of form-function association can be crucial, as it addresses the reliability of the form as a predictor of an interpretation. In other words, the more consistently a form is associated with a given interpretation in language input, the easier learning becomes, because statistical tallying of the association benefits from simpler form-interpretation mapping. On the other hand, if cue-outcome reliability is reduced (as in case of homophony, polysemy and synonymy), learning gets more difficult.

Finally, the third factor, learned attention, takes the focus back to top-down processes related to salience. Learned attention consists of attentional biases, which can depend on different factors, such as L1 interference, redundancy and blocking. In the case of redundancy, linguistic cues that are not essential for the correct interpretation of input are neglected. A typical example is that of grammatical morphemes expressing temporality and appearing together with lexical cues carrying the same meaning (e.g., yesterday he talked), which are reportedly hard to notice and acquire for L2 speakers (VanPatten 2004). Blocking involves a similar process, as an existing and effective stimulus-outcome association blocks further associations of different cues with the same outcome. In sum, learned attention can negatively affect salience of linguistic forms and therefore prevent the subject from registering stimuli. This in turn undermines language acquisition, since attending to stimuli is regarded as crucial for intake, further processing and, therefore, learning.

The link between salience and levels of consciousness is clear, as making a stimulus more or less salient can affect the amount of attention it receives. What is relevant to point out with regard to the awareness debate is that salience is facilitative for learning at both the levels of noticing and detection. Although not implying awareness, detection requires the allocation of cognitive resources in order to register stimuli. Such allocation is not automatic, and it can be boosted by increased salience. In other words, detection can benefit, as noticing does, from the consciousness-raising effects of both perceptual and constructed salience.

1.2.3. Implicit and explicit learning

Salience, the above-discussed models of consciousness in learning and the related detection-noticing debate are the main theoretical notions involved in the definition of implicit and explicit learning. Reber (1967, 1989) defined implicit learning as a process that, without conscious effort to learn, produces from a complex, rule-governed environment a tacit and abstract knowledge which is representative of the structure of the environment and can be implicitly applied and generalized to novel circumstances. This process takes place naturally when the subject is attending to the patterns of variation in the input, without any bias. Similarly, Krashen (1982) claims that acquisition (i.e. implicit learning) is subconscious, the subject only being aware of using the language for communication and not of the process of learning. Williams (2005) states that implicit learning occurs without intention to learn and without awareness of what has been learned. Hulstjin (2005) in his review defines implicit learning as input processing taking place unconsciously, without the intention to find and learn rules and regularities about the linguistic form. Godfroid (2016) adds that implicit learning is apparent in changes in the behavioral responses of the subjects, without the subjects being aware of such changes. Indeed, implicit learning is widely acknowledged as a process that takes place without awareness, even though some researchers do not agree with this position, and question the very existence and effectiveness of implicit learning (DeKeyser 1994; Dienes et al 1991; Dulany et al 1984, 1985; Perruchet & Pacteau 1990, 1991). Conversely, explicit learning is a conscious effort to learn notions about the language, finding out whether the input presents regularities and voluntarily working out the rules (Hulstjin 2005).

According to these definitions, implicit learning corresponds to Tomlin and Villa's learning through detection and results in implicit knowledge, whereas explicit learning, which implies awareness, is related to Schmidt's noticing and understanding and creates mainly explicit knowledge. Despite fairly widespread agreement about the definitions, a strong debate exists regarding the relative roles implicit and explicit learning should play in language pedagogy. There is no consensus on whether or not implicit learning is the only means of second language acquisition, especially for adults, and even among the many researchers who outright reject such a possibility, the role and optimal proportion of explicit learning is still debated.

With his Input Hypothesis and Monitor Hypothesis, Krashen (1982) denies any role for explicit learning in language acquisition. Assuming that adults can deploy the same natural "language acquisition device" children use for their L1, Krashen states that the only way to create implicit knowledge is to be exposed to comprehensible input. Following a grammatical syllabus does not contribute to language acquisition because it prevents real communication through the second language. Moreover, given that acquisition follows fixed sequences, an external syllabus is unlikely to provide the right structure at the right moment for all learners, potentially at different developmental stages. The role for language teaching in this model is only to provide comprehensible input to those who cannot get it elsewhere, either because of a low level of proficiency or of their situation (e.g. foreign language students). Even then, explicit instruction can only create a Monitor, whose role is to edit the output produced by implicit knowledge and supply items that are not yet acquired (if certain conditions are met, see section 1.1). These functions, however, are not considered crucial for language acquisition.

Such an extreme position is not dominant in the literature. On the contrary, numerous researchers argue for a positive role for explicit instruction on acquisition, albeit in different proportions and within different models. Robinson (1995), for instance, strongly supports explicit learning. He empirically asserted that implicit learning alone is not more efficient than explicit on either simple or complex rules. Robinson (1997) further claims that implicit knowledge is not actually rule-based (i.e. the result of the abstraction of rules and regularities in the input) but is instead a memory-based knowledge of instances. DeKeyser (2015) with the Skill Acquisition Theory holds that possessing declarative knowledge is a sine qua non condition for the proceduralization and automatization of linguistic knowledge.

Many researchers assign a more indirect role to explicit learning. Paradis (2004) claims that implicit competence is qualitatively different from explicit knowledge; in other words, the language that is perceived is only the surface manifestation of computational procedures completely inaccessible to awareness, which do not correspond to the conscious abstraction of rules we know as grammar. Therefore, what is available for noticing (i.e. the linguistic output) is deeply different from what can be internalized (i.e. implicit competence in using language). Consequently, noticing and explicit instruction cannot have a direct role in the creation of implicit knowledge, because it is not possible to be aware of it. Nevertheless, Paradis assigns a role to explicit learning: explicit knowledge of the surface forms can be useful as a model for practice, which in turn creates implicit competence. In other words, while the learner practices the form under the guidance of explicit knowledge, implicit learning takes place, because the underlying processes can establish themselves. Paradis also proposes an 'Activation Threshold Model' according to which "the activation threshold of an item is lowered each time it is activated" (Paradis 2004 p. 51); thus, appropriately producing a form (possibly focusing on the meaning rather than the form itself) increases the likelihood of acquiring the underlying computational procedures. In addition, in the first phases of L2 acquisition, adults usually rely on explicit knowledge, so it is reasonable to provide them with correct information; during L2 development, a gradual shift occurs from using metalinguistic knowledge to using implicit competence. Reber (1989) argues for an analogous role for explicit knowledge. Although maintaining that it cannot turn directly into implicit knowledge, he claims that explicit instruction may be useful as long as it increases the salience of the form, thus augmenting the effectiveness of the learner's attentional focus. In other words, instruction can orient the subject's attention, facilitating the creation of correct coding schemes and thus the creation of implicit knowledge.

N. Ellis' (2002, 2005, 2008, 2011) position is in some respects close to Reber's. According to Ellis' (2002) model, language learning consists of "the gradual strengthening of association between co-occurring elements" (p. 173). Therefore, language knowledge is mainly implicit, statistical and probabilistic, created through unconscious tallying of the item frequencies in the input. Aslin & Newport (2012) corroborate such a model: reviewing numerous empirical studies, they claim that statistical learning is an implicit process taking place through mere exposure to the input and is capable of extracting rules and patterns that can be applied to novel contexts. However, according to Ellis (2005), implicit processes alone may not be enough to trigger learning. According to the Law of Contiguity (Ellis 2001), the subject needs to become aware of the new stimulus a first time for it to subsequently become object of unaware, implicit statistical learning (see also § 3.1)

According to this view, implicit learning consists in priming routines or chunks of representations; this process takes place automatically, and it is carried out by what Ellis calls "zombie agents" - i.e. processors that execute routines without and beyond our conscious control (Ellis 2005). The processing strategies and L1 routines automatically driving these systems make the processing usually fast and efficient in the L1, but may actually hamper the correct processing of the input in the L2, e.g. ignoring relevant clues and/or biasing the subject's attention towards the wrong ones. Here the two core concepts of salience and blocking come into play.

The instructor needs to keep into account that psychophysical salience of important linguistic constructions is often so low that they can go undetected and unnoticed. For instance, crucial form-function relationships are often non-salient in the language stream. This fact is true across languages as it responds to the known least-effort principle (Zipf 1949). According to this law, the more frequently a form is used, the more speakers abbreviate it in the attempt to minimize articulatory effort. This leads to the shortening of the most frequent words in languages, which consist in crucial items such as grammatical-function words and bound inflections. Such words become short and therefore non-salient, i.e. hard to perceive in the input and thus unlikely to be learnt without a pedagogic intervention.

The issue of lack of salience does not concern grammar learning only. Formulaic sequences in general and idioms in particular present a similar problem when it comes to L2 learning. Looking at the aspects of perceptual salience already taken into account (perceptual prominence, surprise, privileged meaning) in relation to idioms, it becomes clear that they are likely to have a low salience. Indeed, formulaic sequences and idioms are often composed of frequent words already known to the learners. Therefore, there is little chance for the attention of the speaker to be drawn to the formulaicity of the expression, since its components are nor prominent, nor unknown, nor surprising, and they already present effective form-meaning connections in the speakers' mental lexicon (Boers & Lindstromberg 2009). This is the reason why the learning of formulaic language is likely to highly benefit from constructed salience.

In regards to blocking, experimental studies (e.g. Cintròn-Valentìn & Ellis 2015) demonstrate that in the presence of redundancy (e.g. a lexical and a morphological cue both indicating when an action takes place), the subject gets the communicative content from the most salient cue (usually the lexical one), and this blocks his/her attention from the less salient (usually the morphological one), thus preventing the noticing and the learning of the less salient feature. The phenomenon of blocking is based on an associative process, which shifts attention as a result of previous experience: when the learner implicitly knows that a stimulus is associated with a certain outcome, it becomes harder to learn a different stimulus for the same outcome. The previous experience driving such an attentional bias can either be related to salience (as in Cintròn-Valentin & Ellis' 2015 above-mentioned example), processing strategies (Lee & Benati 2007; Van Patten 2004) or the L1. One example of the latter case is phonology: L1 phonetic prototypes distort the perception of items in order to make them seem more similar to the prototypes themselves.

These phenomena clearly indicate that implicit learning alone can be misleading and flawed, and therefore that explicit learning indeed has a crucial role. This implies that language teaching is necessary and fundamental in order to direct the processing strategies of language learners (Benati 2005). According to Ellis, there are some important actions the language instructor should take. First, teachers should provide explicit instruction about non-salient structures, e.g. by means of constructed salience; this triggers the noticing of the new forms or, in Ellis' words, "seeds" it (Ellis 2005, p. 320). Once the construction is seeded in explicit memory, then implicit learning through statistical tallying can take place. In order to promote and speed up statistical learning, pedagogy should expose the subjects to exaggerated input and provide error-free learning through corrective feedback. To sum up, in Ellis' system, noticing needs to be directed by instruction through the manipulation of salience. In this way, the process of implicitly detecting, and thus learning, new structures in the input can take place.

In addition to the theoretical issues treated so far, one more aspect needs to be considered in regard to the relative roles of implicit and explicit learning of an L2. One of the main reasons the acquisition of a second language does not take place efficiently through implicit learning alone, as L1 acquisition does instead, is related to so-called critical periods. It has been demonstrated that children's capacity for instance learning (i.e. picking up form-meaning incidentally) and, more generally, for implicit learning begins to weaken around puberty, at the same time the capacity for explicit learning gets stronger (Granena 2016; Granena & Long 2013; Long 2017). Although implicit learning does remain an option for adults, it is unlikely to be the sole means of successful and accurate acquisition of a second language after puberty. As Long (2017) maintains, a clear evidence for this claim comes from vocabulary. Nation (2006; 2014) has shown that a NNS needs to know approximately 9000 word families in order to understand newspapers and novels and 6000 to watch videos in the L2. Given adults' reduced abilities for instance learning, it is unviable to learn such a volume of information implicitly, because this would require an amount of time and input incompatible with language courses. However, the very same time constraints related to courses and the huge volume of vocabulary items also preclude the possibility to learn them explicitly, i.e. by teaching them one by one.

These reflections, corroborated by the theoretical claims of the scholars cited above, show that the aim of pedagogic action should be combining explicit and implicit learning. With regard to the relative proportion of the two kinds of learning, three factors in particular among those treated above are taken into consideration. First, implicit knowledge is the priority for language teaching. Second, adults can create it while focused on meaningful practice of the language. Third, an excessive proportion of explicit instruction diverts the learners' attention from meaning to form, thus hampering the creation of implicit competence. These considerations result it the need to identify "the least interventionist, but still effective, forms of instruction" (Long 2017, p. 36). In other words, the aim for pedagogy should be to trigger implicit learning and improve and direct it through a limited proportion of explicit learning.

However, a further issue must be addressed: while explicit learning is a surface process that can be directly driven, this is not the case for implicit learning. Indeed, the claims reported above describe implicit learning as an internal, unconscious process beyond the instructor's control. Therefore, from a pedagogic point of view, it is more meaningful to deal with a different, albeit related, concept: that of incidental learning, which may or may not result in implicit learning and, thus, implicit knowledge. The following section (1.3) focuses on the differences between incidental and intentional learning and the relationship between incidental and implicit learning.

1.3. Incidental learning is the means

The notion of incidental learning is methodologically implied in many studies concerning implicit learning and knowledge. Indeed, Hulstijn (2003) pointed out that many studies that are theoretically concerned with implicit learning actually deal with incidental learning on a methodological level. Nevertheless, the term lacks theoretical interpretation (Hulstijn 2003) and a widely agreed definition (Gass 1999).

As Hulstijn (2003) argues, the notion of incidental learning has been loosely interpreted, and therefore it is theoretically weak. According to Schmidt (1994), three definitions are possible. A first, general interpretation only describes incidental learning in negative terms: learning without the intention to learn. This definition can be found in Hulstijn (1996) and Ellis (1994), and the latter also points out that in this case, consciousness is to be understood as intentionality, i.e. consciously paying attention to what is learned (the incidental learning issue, Schmidt 1990). A second possible definition refers to learning an aspect of the input while paying attention to a different aspect of the same input, i.e. learning something when the conscious goal is to do something else. A classic example is learning new vocabulary while focusing on the general meaning of a text. Long (2017) adopts this definition, claiming that during incidental learning, subjects learn at least part of what they learn without intention to, while their attention is focused on something else. In this regard, Bisson and colleagues (2014) talk of learning happening as a byproduct of another task. A third and more specific interpretation states that incidental learning consists of learning formal features while paying attention to semantic features; for instance, learning a grammatical structure while engaged in meaning-focused communication.

Although incidental learning itself has only been loosely interpreted, its distinction from intentional and implicit learning has been explored extensively. Hulstijn (2003) effectively sums up the main difference between incidental and intentional learning, stating that "attention is deliberately directed to committing new information to memory in the case of intentional learning, whereas the involvement of attention is not deliberately geared toward an articulated learning goal in the case of incidental learning" (p. 361).

A main point in the discussion about incidental learning is its relation to, and distinction from, implicit learning. Although unintentional learning is often considered equivalent to implicit learning, this equation is methodologically and theoretically improper. As Godfroid (2016) maintains, a lack of intentionality is necessary but not sufficient for learning to be implicit; indeed, in order to be involved in implicit learning the subject must also be unaware of what is learned at the point of learning it. A related position is taken by Hulstijn (2003), who claims that implicit learning entails more than incidental learning; quoting Paradis (1994), he points out that being incidentally acquired is only one of the features that define implicit competence, which should also be stored implicitly and used automatically.

This issue is closely related to a methodological concern pointed out by Gass (1999) in her discussion of incidental learning of vocabulary: when talking about incidental learning and, thus, about the role of attention, it is essential to take into account that "pedagogically-induced attention may or may not fit in with the learner attention" (p. 321). In other words, the mere fact that the teacher did not explicitly teach a grammatical form or a vocabulary item does not necessarily mean that the learners do not intentionally focus their attention on it. This crucial distinction between teaching-induced attention and actual learner attention reveals a potentially deep flaw in the methodology of empirical studies of incidental (and implicit) learning, which lack a means to prevent a learning process factorized as incidental from actually taking place intentionally. As Bisson and colleagues (2014) point out, an incidental learning condition has been defined simply as a situation where subjects were not explicitly asked to learn, i.e. they were not informed of the subsequent tests (e.g. Pellicer-Sánchez & Schmitt 2010; Williams 2010). However, this kind of design does not guarantee the actual focus of the learners' attention, especially in the case of studies carried out in classroom settings, where learning is generally the implied goal for any activity. In this respect, Bruton et al. (2011) argue that even when participants are not induced to consciously attend to lexical items during reading, they might actually do so, with within- and between-participants variation that are beyond the researcher's control.

To sum up, pedagogicly-induced incidental learning is the most likely route to implicit learning and thus to the highest goal of language teaching: implicit knowledge. However, it is crucial to take into consideration that an incidental-learning condition does not necessarily entail that implicit learning actually takes place.

Where research methodology is concerned, the literature thus shows two main gaps, on both the process and the product level. From a process point of view, research should be able to verify whether incidental conditions actually result in incidental learning, i.e. experimental designs should be capable of controlling and factorizing the subjects' attention during exposure the input. On the product level, Rebuschat (2013) clearly makes the point that "a significant body of work has focused on incidental learning, […], but these studies do not assess whether the acquired knowledge is implicit" (p. 598). It is clear that data produced by empirical studies that do not take these issues into consideration cannot contribute significantly to the investigation of the effectiveness of incidental learning for the creation of implicit knowledge.

1.4. Summary

This first chapter provides the theoretical basis for the line of argument leading to the design of the present study.

As a starting point, it is claimed that the priority aim for language instruction should be the acquisition of implicit knowledge, due to its desirable features related to online language use. This goal defined, the issue clearly follows of how to induce implicit knowledge gains in language learners. Since explicit knowledge cannot be directly transformed into implicit knowledge, there is the need to involve the learners in implicit learning, i.e. learning without awareness. However, this is problematic under different respects: first, there is little agreement in the literature about the actual possibility for learning to take place without awareness (at the detection level); second, learning without any instruction is prone to salience and blocking issues which are likely to hamper the learning process; third, implicit learning is an internal process and it is not possible for the instructor to have a direct control on it.

The present study aims to overcome the issues related to pure implicit learning. The first and the second problem are dealt with by providing the learners with a form of instruction designed to be as little interventionist as possible: setting incidental learning conditions and manipulating perceptual salience and frequency (§ 3).

The third issue is methodological in nature, and it is addressed by measuring both the participants' level of consciousness at the point of learning and the implicit knowledge gained (§ 4).

In order to deal with the findings, gaps and method of the existing empirical literature in a more focused way, it is necessary to first define and justify the target structure for the present work. Formulaic sequences are thus the object of the next chapter.

Chapter Two: Incidental Learning of Formulaic Sequences

The first section of the chapter aims to motivate and contextualize the importance and difficulty of learning L2 formulaic sequences. Mastering formulaic sequences is crucial for a successful use of a second language. This can be seen and confirmed from a number of different perspectives, which can be organized through the distinction between speaker-external and speaker-internal approaches. At the same time, mastering formulaic sequences in a second language is extremely difficult, even for advanced learners, as the patterns of L2 reception, production and acquisition show. Given such features of difficulty and importance, numerous studies have been carried out in order to verify the effectiveness of different pedagogic techniques for the learning and acquisition of formulaic sequences. The main interventions addressed in the literature have been categorized on the basis of being meant to engage either intentional or incidental learning. The second section discusses the main findings reported in this regard and refers to Chapter 1 in motivating the present study's focus on incidental learning conditions.

Narrowing down to incidental learning, the third section describes how statistical learning takes place in first language acquisition of formulaic sequences and reports empirical evidence of the possibility for L2 learners of benefitting from the same processes.

2.1. Importance and difficulty of formulaic language

More than 40 different terms employed to designate and define formulaicity can be counted in the literature (Wray 2002). Such a variety mirrors a vast diversity of approaches and research aims relating to this area of language, which has been investigated in different branches of linguistics, such as formal-linguistics, pragmatics, corpus-linguistics and psycholinguistics.

For the purposes of the present dissertation, it is useful to organize the variety of possible perspectives through the distinction between speaker-external and speaker-internal approaches (Myles&Cordier 2016, Wray 2008). Speaker-external perspectives describe formulaicity as a language phenomenon, independent of the speaker. Therefore, in this framework factors such as frequency, structural properties and pragmatic functions are taken into consideration in order to define formulaic language. On the other hand, speaker-internal approaches provide a psycholinguistic view and focus on the status of formulaic sequences in the mental representations of the speaker. In other words, they look at whether an individual sequence is actually formulaic or not for a specific speaker, regardless of the formal features of the sequence itself.

Such a distinction is adopted here because it allows attainment of two main aims. First, it shows that the importance of formulaicity is confirmed from a variety of perspectives. Second, it leads to a key statement: whereas speaker-external and speaker-internal definitions usually coincide for native speakers, this is not the case for nonnative speakers. This provides a clear perspective on the difficulties L2 learners experience, while at the same time pointing to the ultimate aim of instruction.

2.1.1. Speaker-external perspectives

A main factor determining the importance of formulaicity is its ubiquity in the language. This area is investigated through a typical speaker-external approach, namely the corpus linguistics' statistical view of formulaicity. By this definition, if two or more words co-occur more often than their individual frequency would predict, then they constitute a formulaic sequence (Durrant & Doherty 2010, Jones & Sinclair 1974). Statistical association scores such as Mutual Information, Log-likelihood and Delta P can be calculated from language corpora and combined with frequency and dispersion in order to objectively measure co-occurrence (Gries & Ellis 2015). Most studies of formulaicity in both first and second language, including the present one, conduct this kind of analysis in order to select the target items and take these measures into account in their data analysis. On a large scale, statistical analysis of corpora demonstrates that formulaicity constitutes about 50% of the language (Conklin & Schmitt 2012, Erman & Warren 2000, Oppenheim 2000). This shows how important it is for language learners to master it in order to gain a proper and nativelike language use. At the same time, dispersion analyses demonstrate that despite the ubiquity of formulaicity, single sequences can have very low frequency of occurrence, and, therefore, it may be unlikely for a learner to encounter them often enough to learn them statistically. These considerations offer a possible explanation for the difficulties L2 learners encounter with formulaic language.

Another speaker-external approach, i.e. the structural perspective, confirms from a different point of view the difficulty of formulaic language for second language learners. From the structural perspective, two main formal criteria are capable of identifying formulaic sequences: non-compositionality and fixedness (Read & Nation 2004). Non-compositionality implies the impossibility of interpreting the sequence literally, as the simple sum of its components. In other words, "the meaning of the expression is not (totally) predictable from its form" (Pawley & Sider 1983: 209). Fixedness refers to the degree to which the sequence can be modified, with respect to word order, inflection, and changes or additions of words. Both non-compositionality and fixedness are considered as continua, and each formulaic sequence in a language can be placed at different points of these continua. Learning the position of formulaic sequences on the continua clearly poses a hard challenge for L2 speakers.

The analysis of pragmatic functions of formulaicity provides further support for the claims about its importance and difficulty. Besides contributing to fluency, formulaic sequences perform crucial communicative functions, partly in a similar fashion to content words, e.g. with referential and ideational functions (Boers & Lindstromberg 2012), partly with different roles, such as organizing discourse, aiding smooth social interactions, or conveying an evaluative stance. Moreover, formulaic sequences are often associated with particular speech events and therefore constitute social conventions, i.e. "conventional label for conventional concept, a culturally standardized designation (term) for a socially recognized concept category" (Pawley & Sider, 1983: 209). This highlights another level of difficulty associated with the use of formulaic sequences: besides the lack of semantic transparency determined by non-compositionality, there is a lack of pragmatic transparency, which refers to the need for knowledge of the social and communicative context formulaic expressions are associated with and are appropriate in. In this regard, Pawley and Syder, in their seminal 1983 paper, introduce the concept of "nativelike selection". According to their position, only a small subset of the possible grammatical sentences are actually nativelike in form, that is, would be judged as ordinary and natural by a native speaker, in contrast to expressions that albeit grammatical, would be perceived as odd or foreignisms (Christiansen & Arnon 2017). For example, "I want to marry you" and "I wish to be wedded to you" are both grammatical sentences and bear the same meaning, but no L1 speaker would use the latter or consider it normal and nativelike.

A wider and deeper perspective on the social role of formulaicity is provided by Wray (2002), when affirming that beyond the pragmatic functions so far listed, formulaic sequences are more capable than other aspects of the language of signaling the speaker's identity as an individual or as a member of a group, as they aid the hearer's comprehension referring to culturally determined concepts. Along the same line, she maintains that formulaic sequences are not only the solution to linguistic problems such as speaking fluently, organizing text or helping the hearer understand. Instead, they constitute a linguistic solution to a non-linguistic problem: promoting the speaker's interests, i.e. the speaker's promotion of self. Those interests include being taken seriously, having physical and emotional needs met, being perceived as important and as a full member of groups. In other words, formulaicity has a crucial role in achieving the very objectives of communication itself. Such a perspective provides a deeper view of the importance of this area of the language for L2 learners.

The speaker-external perspectives so far reviewed also allows the analysis and classification of the various expressions of formulaicity a language presents. Numerous taxonomies have been proposed, based on either theoretical or practical needs. Theoretical classifications refer to different models of language, while practically motivated taxonomies are driven by their specific purpose and the by the feature(s) (e.g. internal structure, function, form, meaning) chosen as pivotal (Wray 2002).

For the purpose of the present dissertation, the following distinction is adopted, which identifies main categories of formulaic sequences based on meaning (a continuum from totally idiomatic to totally literal meaning) and function.

- Exclamations (e.g. what the heck). Connotative and non-compositional, they have the role of conveying an evaluative stance.

- Idioms (e.g. kick the bucket). Similarly to exclamations, idioms are characterized by a non-literal meaning.
- Pragmatic formulae (e.g. see you later). Literal meaning, they provide smooth social interactions.
- Function words (e.g. having said that). Literal meaning, they contribute to organizing the discourse.
- Irreversible binomials (e.g. black and white).
- Complex verbs (e.g. talk it over). Literal meaning, referential and ideational function.
- Collocations (e.g. declare war). Literal meaning, referential and ideational function.
- Lexical bundles (e.g. one of the). The only hallmark is higher than chance frequency.

(Boers&Listromberg 2012, Siyanova-Chanturia & Van Lancker Sidtis, 2019).

2.1.2. Speaker-internal perspective

To sum up, statistical, pragmatic and structural analyses focus on the language rather than the speaker, and they are capable of showing the ubiquity, importance and difficulty of formulaicity, at the same time being a basis for possible classifications. On the other hand, speaker-internal approaches investigate the psycholinguistic status of formulaicity, i.e. they consider a sequence formulaic if a speaker retrieves it with greater efficiency, experiencing a processing advantage. This implies that any particular string may or may not be formulaic for any particular person, depending on speaker-related factors, such as linguistic environment, proficiency, communication needs and social context, and irrespective of statistical, structural and pragmatic features of the word sequence itself. The most widely-used definition of formulaic sequence from a speaker-internal or psycholinguistic perspective is provided by Wray (2002):

> "A sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar." (p. 9)

Wray's definition has much in common with one of the first discussion of the FL phenomenon, i.e. Sinclair's Idiom Principle (Sinclair 1991):

> "The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments." (p. 110)

Indeed, both definitions assume that formulaic strings reflect holistic storage. However, while numerous studies show shorter reading times and therefore a processing advantage for formulaic vs. non-formulaic phrases, very few experiments have a design capable of proving assertions about the speakers' internal linguistic representations (Siyanova-Chanturia & Martinez 2014; Siyanova-Chanturia 2015). In other words, the methodology the large majority of applied linguistic studies adopt in order to investigate online processing does not probe deeply enough to provide psycholinguistic information about the kind of storage the speaker's brain employs. The few reliable studies investigating this issue do not actually confirm a holistic storage hypothesis. Arnon and Snider (2010) measured reaction times to 4-word strings, which varied in frequency while having the single components' frequency controlled for. The analysis shows that higher frequency phrases were responded to faster, and the effect was a gradient one found across a wide frequency range. Interestingly, the continuous frequency of the strings was a significant predictor of reaction times, while a binary measure (high-frequency vs. low-frequency combinations) was not. Therefore, despite not being focused on the representation issue, the study argues against a holistic storage of the most frequent sequences (formulaic), opposed to the computation of less frequent combinations (non-formulaic). In a subsequent study investigating representations of multi-word units, Arnon and Cohen Priva (2013) manipulated frequency of strings and their components, looking in this case at the duration of words in naturally elicited speech. According to their hypothesis, if the sequences were stored holistically, then the effects of single-word-related factors such as frequency should not be perceivable. On the other hand, if frequency of the single words significantly affected a formulaic sequence, then this would point to a single-word-retrieving process, and therefore against holistic storage. Results from this study showed a diminished but still significant effect of component frequency, suggesting that formulaicity makes the parts within the whole less salient but still available. Idioms are often considered as the prototypical candidates for holistic storage (Pinker 1999), but even studies in this area argue against such representational status.

Other studies, too, have addressed idiom processing and representations, rejecting the idea of a holistic, word-like storage. Sprenger, Levelt and Kempen (2006) carried out three priming experiments, involving participants in (i) production of idioms and literal phrases, (ii) idiom completion and (iii) asking them to switch between idiom completion and naming. Results supported the hypothesis of separate access to the single words constituting the idiomatic expression, with decomposition involved both at a semantic and phonological level. Such outcomes are reported to support a hybrid account of idiom representation, i.e. the assumption of idioms being both unitary and compositional, at different levels of their cognitive representation. Konopka and Bock (2009) employed a priming-based design, as well, but focused on the syntax level. Their second and third experiments are of special interest here. They compared the syntactic priming ability of idiomatic and non-idiomatic phrasal verbs, and showed little difference between the two. Such an outcome indicates that the magnitude of structural persistence is not predicted by variation in idiomaticity, due to the same need for an abstract phrasal frame for both idioms and non-idioms. Therefore, Konopka and Bock's (2009) results are not compatible with the hypothesis of a lexical, holistic storage of idioms. Tabossi, Fanari and Wolf's (2009) study is worth mentioning here because it is one of the very few working with Italian idiomatic expressions. The study design was based on comparing how long it took L1 speakers to judge the acceptability of idioms, clichés and their matched controls. Results showed that the fact idioms are recognized faster is not related to their idiomaticity, but rather to being well-known expressions. In other words, these data add to the evidence against a holistic storage of idioms.

Such findings show that the issue of the mental representations of formulaic language is still controversial. Moreover, it needs to be pointed out that determining whether the phrases are mentally represented as wholes or not is not crucial for the purpose of the research questions this dissertation seeks to answer, i.e. the possibility for L2 speakers to gain the same processing advantage from the use of formulaicity as natives. Indeed, the strong implications for the way language is noticed, processed and learned derive from the sensitivity to frequency and probability distributions at the phrase level, and this does not necessarily involve holistic storage (Siyanova-Chanturia 2015). Therefore, the choice is made here to adopt a speaker-internal definition which focuses on processing rather than on representations, provided by Myles and Cordier (2016):

> "A psycholinguistic FS is a multiword semantic/functional unit that presents a processing advantage for a given speaker, either because it is stored whole in their lexicon or because it is highly automatised." (p. 10)

The main point is that speakers benefit from formulaicity because relying on ready-made sequences instead of creating new ones online reduces memory load and allows them to save cognitive resources for other kinds of concomitant activities. As Pawley and Sider (1983) point out, human brains possess huge memory capacity, while they are not so good at performing many mental acts at the same time. Therefore, using holistic or automatic units as building blocks of the discourse looks like the most natural, obvious, economical and efficient choice for the way the human brain works. This statement confirms and also explains the ubiquity and the importance of formulaic language pointed out by the speaker-external analyses.

The processing advantage of formulaicity, theorized by Pawley and Sider and included in Wray's and Myles and Cordier's definitions, has been demonstrated in empirical research about L1 speakers. Studies investigating processing of formulaic sequences share the need to employ online measurements, the most common being self-paced reading and eye-tracking. These techniques rely on the assumption that shorter reading times (for self-paced reading) and shorter and fewer fixations (for eye-tracking) imply shorter processing times, and therefore suggest that the speaker is dealing with the language in a more economical and efficient way. Other online measurements employed in less recent empirical studies include timed grammaticality judgments and timed recognition tests, both pc-supported.

Tabossi, Fanari and Wolf (2009) demonstrated through an online judgment test for Italian native speakers that known idiomatic expressions are recognized faster than their controls. A similar test was employed by Arnon and Snider (2010), who addressed the effects of 4-word string frequency on processing latencies of L1 speakers. Their results demonstrate phrase frequency to be a predictor of reaction times, showing an advantage in how native speakers process formulaic sequences.

Underwood, Schmitt and Galpin (2004) carried out one of the first eye-tracking studies, investigating native and nonnative processing of formulaic and non-formulaic phrases. L1 speakers' fixations were shorter and fewer for the last word of formulaic expressions than for control words embedded in non-formulaic phrases, which shows more efficient processing. Along the same line are results from Siyanova-Chanturia, Conklin and Schmitt (2011), who monitored with eye-tracking reading patterns of idioms and novel phrases in native and nonnative speakers. L1 speakers' data confirmed the processing advantage expected. These results were confirmed in a more recent eye-tracking study of idioms (Carrol, Conklin & Gyllstad 2016). Siyanova-Chanturia, Conklin and Van Heuven (2011) worked on a different typology of formulaic sequences, i.e. binomials and their reversed forms, and the eye movements of native speakers in this case also confirmed a processing advantage for the formulaic version of the phrases.

Self-paced reading studies provide a similar pattern of results to the eye-tracking data. Conklin and Schmitt (2008) had native and high-proficiency nonnative speakers read both formulaic sequences and matched non-formulaic phrases. L1 speakers' results showed shorter reading times for formulaic sequences than for control phrases, thus proving that formulaic sequences are indeed processed more easily than non-formulaic controls. A different methodology was employed by Jiang and Nekrasova (2007), where native and nonnative speakers judged the grammaticality of formulaic and non-formulaic phrases in a timed task. Results confirmed a processing advantage for formulaic sequences in L1 speakers.

Summing up, empirical studies have employed various types of online measurements in order to investigate the way L1 speakers deal with different kinds of multi-word sequences. All report converging results, showing that NSs show greater efficiency when processing sequences that external criteria such as statistics identify as formulaic. This implies that for native speakers the internal and external definitions of formulaicity tend to coincide. What is noteworthy is that this is not always the case for nonnative speakers, whose processing of formulaic sequences may or may not be economical, depending on such factors as proficiency and exposure. Indeed, in the same experiments that demonstrate a processing advantage for L1 speakers, results are mixed when it comes to L2 speakers.

Siyanova-Chanturia, Conklin and Schmitt (2011), who found more efficiency in L1 speakers' eye movements for idioms than for novel phrases, report no difference in the processing patterns of L2 speakers between formulaic and non-formulaic language.

Other studies suggest that nonnative speakers might benefit from formulaicity as L1 speakers do, albeit individual features of both the speaker and phrase seem to play a crucial role, unreported in the L1 results. Conklin and Schmitt's (2008) self-paced reading results show that high-proficiency L2 speakers read formulaic sequences faster than control phrases, thus confirming the possibility for NNSs to develop processing advantages for formulaicity, as native speakers do. Similarly, in Jiang and Nekrasova (2007), L2 learners show a comparable pattern to L1 speakers, providing faster and more accurate judgments for formulaic than for non-formulaic sequences.

As for eye-tracking studies, Underwood, Schmitt and Galpin (2004)'s advanced nonnative speakers show longer and more numerous fixations overall than native speakers, but at the same time, both L1 and L2 speakers fixate on the last word in formulaic expressions less than control words embedded in non-formulaic phrases, thus showing similar processing patterns. Despite the noteworthy results, the often-cited study presents some methodological limitations that need to be taken into account. Namely, Underwood and colleagues did not control for the actual knowledge L2 speakers had of the target formulaic sequences, and instead admit that a post-hoc partial investigation revealed that the expressions were mainly unknown. Therefore, attributing the results of the eye-tracking measurements to automatic processing of the formulaic sequences is clearly problematic.

More reliable results are provided by other eye-tracking studies, such as Carrol, Conklin and Gyllstad (2016), which focused on the interaction between L1 and L2 when reading L2 idioms. The processing advantage found for L1 speakers was confirmed in nonnatives, too, with the authors affirming "there is nothing fundamentally stopping L2 speakers from instantiating idioms in the mental lexicon in a way that enables them to process them quickly, in the same way as native speakers" (p. 433). However, the discussion pointed out the crucial role of proficiency and exposure for this process to take place.

Siyanova-Chanturia, Conklin and Van Heuven (2011) made a similar point in their study of binomials, which included native speakers, low proficiency and high proficiency nonnative speakers. L1 and highly proficient L2 speakers showed similar sensibility to violations of formulaic sequences, whereas low-proficiency L2 learners had the same reading patterns for both binomials and their reversed forms. These findings were interpreted as highlighting the crucial role of proficiency and frequency of exposure.

More evidence of the effects of frequency comes from a study by Kim and Kim's (2012) of complex verbs. Their self-paced reading, meaning-focused task showed that collocational frequency significantly affected both native and nonnative speakers' reading times. However, while L1 speakers were sensitive to the differences between high-, mid- and low-frequency targets, L2 learners' data only displayed a difference between high and low frequency phrases. The authors claimed that these results showed a holistic storage of formulaic sequences in both native and nonnative speakers. Their design and methodology did not actually allow such a psycholinguistic statement (Siyanova-Chanturia 2015), but they did support the idea that L2 learners can achieve a nativelike statistical sensitivity to co-occurrences.

Moreover, this study was especially effective in showing how speaker-internal and speaker-external definitions of formulaicity coincide for native speakers, whereas they do not - but could - for nonnatives.

In their review of the existing experimental literature about L2 processing of formulaicity, Siyanova-Chanturia & Van Lancker Sidtis (2019) argued for a similar interpretation, observing that self-paced reading studies show overall shorter reading times for more frequent sequences than for less frequent ones, for both L1 and L2 speakers (Hernandez, Costa and Arnon 2016, Trembaly et al 2011).

In sum, what emerges from the empirical literature about nonnative processing of formulaic language is that L2 learners are potentially capable of benefitting from the same processing advantage L1 speakers experience. However, this possibility is conditional on such factors as a subject's proficiency and the sequence frequency. Those factors point to a crucial role of frequency of exposure in the process that allows a sequence to become formulaic from a speaker-internal perspective.

2.1.3. Idioms

A special note is needed when it comes to idioms. On a theoretical level, idiomatic expressions present the same features of difficulty and importance so far identified for formulaic language in general, i.e. they are numerous but dispersed in the language, they present a variable and not predictable degree of fixedness, they are pragmatically opaque and culturally determined. The considerations reported so far about the processing of formulaic language were supported by empirical evidence from both idiomatic and non-idiomatic phrases, because results show similar patterns and can therefore be treated together. However, in the literature, they are often acknowledged as even more problematic than other kinds of FSs (e.g. Macis&Schmitt 2017; Irujo 1986). The highly idiosyncratic semantics that distinguishes idiomatic expressions from literal, compositional language can make a difference and has to be taken into account. One of the possible effects this feature has is a reduced role for frequency and the possibility for idioms to become part of a speaker's repertoire even after brief exposure. Reuterskiöld and Van Lancker Sidtis (2013) verified this hypothesis with L1 children, proving that the unitary, non-compositional and linguistic-contextual characteristics of idioms can make them possible to acquire from a single exposure.

Another line of research has investigated whether the literal or the idiomatic meaning that is activated first and with less effort.

Martinez & Murphy (2011) exposed Brazilian learners of ESL to short texts composed of the top 2000 English words. Half of the texts included formulaic expressions with a high degree of non-compositionality, while in the other half only literal meanings occurred. Comprehension test scores were compared between texts including figurative meanings and texts with literal-only use of words. Results claim not only that the presence of multiword expressions undermined the subjects' comprehension, but also that the participants overestimated their comprehension of the passages. Namely, their knowledge of the single components of the FSs led to misunderstanding and lack of noticing. The L2 speakers' tendency to interpret the idioms literally has also been demonstrated through online, timed lexical decision tests. Cieslicka (2006) exposed the subjects to non-defining sentences containing familiar idioms and then asked them to choose between four lexical items, related to either the figurative or the literal meaning of the target phrase. The semantic associations related to the single words showed to be more powerful than those linked to the figurative meaning of the expression.

Similarly, an eye-tracking study by Siyanova-Chanturia, Conklin and Schmitt (2011) showed longer reading times for figurative than for literal readings. Therefore, as Conklin and Schmitt's (2012) review highlights, "even for highly proficient nonnative speakers, processing may be slowed when idioms are used figuratively" (p. 50). In other words, speaker-external features such as a highly idiosyncratic semantics seems to constitute an additional difficulty for nonnative speakers in dealing with idiomatic expressions.

Studies demonstrating that literal meaning is activated first and dealt with more economically than figurative meaning show an additional factor of difficulty for idioms learning in relation to other kinds of formulaic sequences. At the same time, studies at the electrophysiological level show through event-related brain potentials (ERPs) that both L1 and L2 speakers benefit from the prefabricated and hence highly predictable nature of idiomatic expressions once they are acquired (Moreno, Federmeier & Kutas 2002; Siyanova-Chanturia, Conklin, Caffarra, Kaan & van Heuven 2017; Siyanova Chanturia & Van Lancker Sidtis, 2019). This finding adds to those about formulaic language in general, demonstrating that L2 speakers can gain the same processing benefits as L1 speakers when dealing with idioms.

Theoretical and empirical literature about formulaic language highlights its importance at both a speaker-external (ubiquity in the language, fluency, nativelike speech, pragmatic and communicative essential functions) and speaker-internal (processing advantage) level. At the same time, evidence from L2 processing and use demonstrates how difficult a challenge formulaicity poses to language learners. On a receptive level, the existence of increased processing efficiency is proven only for advanced learners, who have had the opportunity to be exposed to massive amounts of input. At the same time, on a productive level, formulaicity is known to be one of the last areas where L2 learners close the gap on native speakers, and longitudinal studies attest that most NNSs never attain a nativelike level, even when of very high proficiency (Conklin & Schmitt 2008). Durrant and Schmitt (2009) found that L2 learners do make use of very frequent collocations, yet their writing does not sound native-like, due to the lack of low-frequent but strongly-associated forms (e.g. "densely populated", "preconceived notions") which are highly salient for native speakers. In other words, the main tendency among nonnative speakers is to over-use a limited number of chunks they are confident with, creating so called "islands of reliability" (Dechert 1983), or "lexical teddy bears" (Hasselgren 1994), while they are not capable of correctly employing less-frequent yet salient sequences.

Different approaches, including longitudinal studies, psycholinguistics, corpus linguistics and pragmatics, all demonstrate the importance and difficulty of L2 formulaic language acquisition. Scholars and practitioners have devoted considerable effort to experiment with and validate different pedagogic interventions to improve the acquisition process. The following section addresses the main outcomes of the empirical literature, focusing on studies comparing intentional and incidental learning conditions.

2.2. Teaching formulaic language: intentional and incidental learning conditions

A main variable to be taken into account when investigating pedagogic techniques capable of improving learning and acquisition of formulaic language is the degree of intentionality involved in the process. In other words, there is agreement in the literature on a distinction between learning conditions meant to engage the speaker in intentional or deliberate learning, on the one hand, and incidental learning on the other. Learning conditions are considered to be intentional when the student is forewarned that he/she is going to be tested about vocabulary and/or when instructions explicitly focus on lexical items (Nation & Webb 2011; Pellicer-Sánchez & Boers 2018). In this context, learners are involved in activities whose explicit aim is for students to learn formulaic sequences, and they are clearly made aware of that aim. Three main strategies have been adopted to design intentional learning conditions (Pellicer-Sánchez & Boers 2018). First, students are explicitly asked to look for FSs while reading texts. Second, they are involved in decontextualized activities meant to intentionally commit FSs to memory, such as studying word lists, using flashcards, or working with vocabulary exercises; these activities are not necessarily related to any text. Third, the instructor highlights special features of FSs capable of fostering memorization, such as a link between literal and figurative meanings, or sound patterns like alliteration, rhyme and assonance.

When it comes to incidental learning conditions, as previously described at length (§ 1.3) the focus remains on the content of linguistic input, so that FL learning takes place as a by-product of such focus. In order to reach that goal, the instructor typically involves learners in extensive reading or reading-while-listening, but also listening alone and multimodal learning conditions (Pellicer-Sánchez et al 2018).

A number of studies have compared the effectiveness of intentional and incidental conditions on the learning of FSs.

Boers, Eyckmans, Kappel, Stengers and Demecheler (2006) carried out a small-scale study to assess the effects of explicitly drawing the learners' attention to FL. Two intact classes for a total of 32 L2 English speakers were randomly assigned to either the experimental or the control group, both exposed to the same authentic reading and listening material for 22 teaching hours. In the experimental group, the teacher emphasized the syntagmatic dimension of vocabulary, and students were encouraged to notice collocations and idiomatic expressions. The control group was taught by the same teacher for an equal amount of time with identical material, but no focus on the formulaic nature of language was provided. The two groups were then interviewed to assess their oral proficiency and use of formulaic sequences. The interview consisted of two parts: a conversation about a short text they were required to read and free production. Blind judges unaware of the group division provided the proficiency scores and the formulaic sequences counts. The scores showed the experimental groups to be significantly more proficient than the control group, and to use significantly more formulaic expressions. However, such results were not due to the experimental-group students producing the formulaic sequences learned during the treatment, but to their greater capacity to re-use the collocations found in the prompt text they read before the interview. Confining the FSs counts to the free-production part of the interview, no difference emerged between the experimental and the control group. In other words, the study findings highlighted a strategic advantage of the experimental-group subjects, but no actual learning of the FSs they encountered, at least at the level of production.

Peters ran two studies (2009, 2012) where one of the variables was again the explicit encouragement to focus on FSs when reading texts. In the 2009 paper, the sample was randomly divided into two groups, exposed to the same reading text, where vocabulary items and collocations were typographically enhanced and glossed. One of the groups was encouraged to focus on new vocabulary, while the other was explicitly instructed to pay attention to the collocations. Unlike the Boers et al study, the posttests here targeted the items in the text. The scores showed significant learning effects for both groups, and no difference due to the request to focus on FSs. The 2012 experiment is a conceptual replication of the earlier study, with L2 German instead of L2 English, and an extra variable added by not enhancing some of the glossed items. Again, enhancement proved to be effective, while the explicit encouragement to focus on collocations had no effect on the knowledge gained. It is worth mentioning that all subjects were alerted to the fact that a vocabulary test would have followed the reading activity. Therefore, it might be problematic to consider the learning conditions subjects were exposed to as incidental, even for participants not explicitly asked to pay more attention to FSs.

More closely controlled in this respect was the 2012 study by Szudarski. Forty-three intermediate L2 learners from three intact classes were assigned to two experimental groups and a control group. The experimental groups were exposed to the same reading materials containing the target collocations over a period of three weeks. While the first group additionally carried out explicit exercises focusing on collocations, the second received no mention of the FSs embedded in the texts. Posttests measuring both receptive and productive knowledge showed that the explicit-instruction group significantly outperformed the incidental-condition group.

Similarly, Laufer and Girsai (2008) assigned three intact classes to one of three experimental conditions, all of which included reading the same text containing the target collocations. In the following lesson, the three groups were respectively involved in (i) meaning-focused activities, (ii) vocabulary fill-in-the-gap and multiple-choices exercises and (iii) contrastive analysis and translation exercises. Passive and active recall posttests highlighted that the explicit-exercise group learned more new items than the incidental-condition group, and that the contrastive-analysis group outperformed both of them.

All the studies reviewed so far adopted the first among the three strategies mentioned above to create intentional learning conditions, i.e. they focused the learners' attention on the FSs occurring in a reading text. In contrast, Sonbul and Schmitt (2013) assessed the effectiveness of the second strategy, namely decontextualized exercises, comparing it to incidental-learning conditions. Their 43 subjects were randomly assigned to three groups. The first was presented with a reading passage in which the target collocations were embedded. The second group read the same passage, but the collocations were typographically enhanced. Finally, the third group saw the target items in isolation and was required to memorize them. Both immediate and delayed posttests showed that the decontextualized and enhancement conditions resulted in new knowledge, with no significant differences between them, and that both brought about more learning gains than the purely incidental condition.

In sum, empirical research provides mixed results, with some evidence showing a stronger learning effect for intentional learning conditions involving elaboration, i.e. a mental operation focusing on the form of the words (Barcroft 2002; Boers & Lindstromberg 2012). However, when evaluating the effectiveness of incidental learning conditions, it needs to be remembered that extensive reading and/or listening not only affect vocabulary and FL learning, but also involve the development of broader linguistic competences not implicated in decontextualized memorization activities. Therefore, adopting incidental learning conditions allows the instructor to optimize available time, simultaneously fostering different levels of linguistic competence.

On a different level, incidental learning conditions might be preferable because the focus on content, not on language form, makes them more capable of leading to implicit rather than explicit knowledge (§ 1.3). In this regard, it is crucial to point out that an incidental, statistical approach to the teaching of formulaic language is especially appropriate also because its features make it more likely to be stored implicitly in the speakers' mind than other lexical aspects of the language. Referring to Nation's (1990, 2001) categorization of the different dimensions of vocabulary knowledge, Ellis (1994) argued that a distinction is possible between two groups of components: (i) aspects that are related to form-meaning connections and can be better learned explicitly; (ii) aspects that concern form, are usage related and can be learned implicitly. Collocational knowledge is considered as being part of this second group (R. Ellis 2004).

These considerations add to the motivations for privileging incidental learning conditions for formulaic language teaching, and thus aiming at the creation of implicit knowledge.

The process likely leading to this outcome and the main factors involved are treated in the next section.

2.3. Incidental learning of formulaic sequences in first and second language

According to the Law of Contiguity (Ellis 2001), two vocabulary items often co-occurring in the input become associated in long-term memory, and this happens largely through implicit processes. In other words, long-term memory is statistically sensitive to frequent collocations, so that when the same items are encountered in subsequent input, they can benefit from a processing advantage. At the first encounter with a FS, an association is consciously made, and then the collocational knowledge is created and stored without the subject being aware of it, through implicit tallying processes affected by frequency. Ellis describes this process of collocation learning as being part of L1 acquisition, and different views exist in the literature about whether it applies to adult L2 learning or not.

Wray (2002) claims that adult learning takes place in a crucially different way, as L2 adult learners break up collocations and learn the words separately, retaining no information about co-occurrence. In her view, this happens for a combination of social and cognitive reasons. On a social level, adult L2 learners lack the L1-child pressure which leads them to learn whole helpful communicative sequences. On the contrary, traditional classroom teaching methods are more likely to focus on form and single new words. On the cognitive side, the fact of being literate and therefore aware of words as basic units of language makes it uncomfortable for adults not to know how sequences are broken down into component words. Consequently, according to her view learning of collocations only takes place intentionally, with learners noticing a gap in their knowledge and attempting to memorize formulas.

Arnon and Christiansen (2017) consider the diversity of FL acquisition between native and nonnative speakers so deep to give it a crucial role in the difference between L1 and L2 acquisition in general. According to their view, FSs can be acquired as a product of two mechanisms. The first is undersegmentation, i.e. a FS is acquired as a whole and only later is it properly segmented. The authors state that adults are not likely to undergo such a process, due to their knowledge of single words, their metalinguistic awareness of words as the units of language, and the kind of input thy are exposed to, which is less repetitive than children's. The second possible way of acquiring FL in this framework is chunking. As a result of frequent co-occurrence, individual words can activate a phrase and the phrase in turn can activate the individual words. Both adults and children may form FSs in this same way, but given that children learn the L1 word and the concept it refers to at the same time, they are capable of learning predictive relationships from the grammatical elements involved in the phrase. L2 adult learners, conversely, already have conceptual representations, so they simply map novel labels onto existing concepts, thus missing the predictive potential of grammatical elements. As a consequence of the processes described, Arnon and Christiansen maintain that FSs work as building blocks for children in the acquisition of their L1, but not (or to a much lesser extent) for adults learning an L2. It is noteworthy, however, that they consider it possible for co-occurrence frequency to have a role in the chunking process, even for adults.

Usage-based perspectives report a similar idea about adult L2 learners employing collocational information like L1 children. According to Wulff (2019), L2 learners implicitly tally and tune their constructional knowledge according to words that preferably occur together. Similarly to Arnon and Christiansen (2017), Wulff ascribes to FL a further role in both first and second language acquisition. In her view, not only are FSs learned as chunks, but they are also capable of triggering a process called bootstrapping. That process leads learners to deconstruct lexical exemplars learned as wholes, in order to recognize patterns and create generalizable constructions. Crucially, she maintains this process to be available not only for children, but also for adults, which makes frequent FSs potential acquisition kick-starters for L2 learners.

There are two main empirical studies which aimed at collecting experimental data comparing L1 and L2 acquisition of formulaic sequences.

The first study was carried out by Durrant and Schmitt (2010), in order to empirically address the issue of whether there is a qualitative difference between the ways L1 children and L2 adults acquire FL. In a laboratory setting, they exposed experimental subjects to one of three conditions: single exposure, verbatim repetition (i.e. same collocation in the same context) and repeated use of the same collocation in different contexts. The first condition aims to test the first trace a collocation is supposed to leave in the learner's language system, in order for subsequent implicit learning to take place. The second and third conditions address the effects of different kinds of repetitions on learning. Given that strongly associated sequences are often not frequent enough to be statistically learned by L2 speakers, one solution could be for teachers to artificially enrich the input with more occurrences of target collocations. Testing the effects of repetitions is therefore not only relevant to the psycholinguistic study of FL acquisition, but also has direct pedagogic implications. Participants exposed to the three conditions carried out a naming task, where they were shown the first word of the FS (an adjective) and two letters of the second word (a noun) and they were asked to say the missing word aloud. Recall rates were compared between nouns following the same adjective they occurred with in the treatments and nouns paired with different adjectives. Results showed that in all three conditions, nouns were significantly better recalled when following the adjectives they were paired with in the training phase. The size effect was weak for the single-exposure condition and large for both the repetition conditions, with some advantage of the verbatim over the varied repetition. Despite a weak effect size, results showed that adult L2 learners do retain memory of co-occurrence after a single exposure, and do so without a conscious intention to learn. This suggests that NNSs' difficulties in L2 FL mastering are likely due to insufficient exposure to appropriate L2 input, rather than to a word-based approach to learning. Such finding constitutes a counterevidence to Wray's position about intentional learning of FSs broken down into their components.

The second study empirically focusing on this issue is more recent, and especially relevant as it employs online measurement of eye movements (eye-tracking). Yi and colleagues (2017) had among their goals to empirically verify whether usage-based statements about statistical learning apply to both native and nonnative speakers when dealing with FSs. They addressed the issues by measuring statistical sensitivity to FSs frequency and contingency in L1 and L2 Chinese speakers. The experimental subjects read 80 critical sentences containing FSs controlled for frequency and mutual information score, and followed by comprehension questions, while their eye-movements were recorded. Results from both early and late eye-tracking measures (for the difference between early and late eye-tracking measures see § 4.5) showed adult L2 learners retain statistical learning ability, so the data point to the possibility that nonnatives share statistical learning mechanisms with native speakers when processing FSs.

Empirical results such as those reported (Durrant & Schmitt 2010; Yi et al 2017) support the idea that Ellis' paradigm of FL acquisition for L1 speakers apply to L2 learning, as well. This being the case, nonnative speakers could exploit statistical learning of co-occurrences to reach the processing advantages natives benefit from.

2.4. Summary

The aim of the present chapter is to justify the choice of formulaic sequences as a target structure, and to deal with the main claims in the literature about FL teaching and learning in an L2.

The crucial importance of mastering formulaic language to communicate in a second language effectively is recognized and demonstrated from many different perspectives. Namely, speaker-external perspectives (e.g. statistical, pragmatic, structural) confirm the ubiquity and relevance of FSs in a language. On the other hand, the speaker-internal angle shows the psycholinguistic advantages L1 speakers benefit from when using FSs, thus defining a desirable aim for L2 FL acquisition. At the same time, both the study of FSs features and the research on L2-speakers' performance point to formulaicity as one of the most difficult challenges for NNSs. This being the case, in order to foster L2 FL acquisition, instructors and researchers have experimented and investigated different techniques, which are usually categorized on the basis of the kind of learning condition they expose learners to, i.e. either intentional or incidental. Empirical studies comparing the relative effectiveness of intentional and incidental learning have produced mixed results, with intentional learning conditions involving elaboration showing some advantage. However, such studies have measured knowledge by means of offline tests or production, with no data on implicit knowledge gains. The present dissertation aims to fill this void in the literature by providing data from online measures at both the process and the product level, and by focusing on incidental learning, because of the greater likelihood of its leading to the acquisition of implicit knowledge.

The process from incidental learning conditions to implicit knowledge of formulaic sequences is described by Ellis for L1 speakers as the Law of Contiguity. It implies, first, the creation of a trace by means of conscious noticing of the unknown FS, and then the development of implicit knowledge through statistical tallying of subsequent co-occurrences of the words composing the phrase. Empirical research has been carried out to demonstrate that this process may take place for L2 learners, as well.

As a consequence, pedagogic techniques ought to be designed and employed in order to boost such a process. The present study aims to do so by means of enhanced incidental learning, which involves a twofold manipulation of learning conditions. In order to facilitate the first phase (i.e. the creation of a trace in the learner's mind), salience of the first occurrences of the target items is increased by means of unobtrusive input enhancement. Then, frequency is artificially augmented in the context of incidental learning conditions, with the aim of promoting statistical learning.

A deeper reflection on the theoretical and empirical underpinnings of enhanced incidental learning is the subject of the next chapter, which also reports evidence about the components of the pedagogic technique designed: input enhancement, incidental learning conditions and increased frequency.

Chapter Three: Enhanced Incidental Learning

The first chapter provided the theoretical bases for the importance of gaining implicit knowledge. Moreover, it showed that to achieve this aim, learners need to be involved in incidental learning conditions that unobtrusively manipulate salience in order to speed up and direct the learning process while keeping it as implicit as possible. The present chapter describes a pedagogic technique aimed at implementing this strategy in the classroom: enhanced incidental learning. As discussed in the second chapter, the linguistic domain chosen is formulaic language. The first section of the chapter recaps the rationale underlying enhanced incidental learning and explains how it is implemented in the classroom. The proposed technique combines incidental learning conditions with unobtrusive enhancements, in order to manipulate the conditions under which the input is experienced.

The following sections deal with the pedagogic practices involved. The first is input enhancement, a tool the instructor can use to manipulate the learning process and the perceptual salience of linguistic input to different degrees of noticeability. Input enhancement is defined theoretically from the perspectives of different frameworks in § 3.3. Then, the empirical results of experimental studies addressing the effectiveness of four kinds of enhancement devices (including increased frequency) for the learning of grammar, vocabulary, and formulaic language are reported. The analysis of the existing literature highlights some gaps in the research, as well as methodological weaknesses of some previous studies, further justifying the present experiment (§ 3.4).

The last section of the chapter deals with the chosen format for incidental learning conditions. Reading while listening is theoretically framed, and empirical evidence supporting it is reported (§ 3.5).

3.1. Enhanced incidental learning

Chapter 1 argued that the main goal for language instruction should be the acquisition of functional use of the L2, i.e. its automatic access, which makes the language available in the context of online listening and spontaneous production. In other words, L2 speakers need implicit knowledge of the language (§ 1.1).

Despite the debate about the best pedagogic interventions and the possible benefits of explicit instruction, it is safe to state that implicit knowledge is acquired unconsciously, as a byproduct of a focus on meaning and real communication. In terms of language instruction, this means more class time devoted to genuinely communicative tasks and activities, i.e. to the creation of incidental learning conditions, where students can be exposed to the language and use it while focusing on content. However, psycholinguistic and pedagogic research demonstrates that the amount of input a language course is able to provide is not sufficient for incidental learning to take place effectively and significantly.

Therefore, instructors must speed up the process while at the same time maintaining the aim of creating implicit knowledge, that is keeping learning conditions as incidental and implicit as possible.

Focus on form (e.g., Long 2015) has been proposed as a solution to this issue. Indeed, a temporary, reactive, and in-context switch of the focus from meaning to form is a reasonable and robust means of facilitating noticing and speeding up the learning process., considering the nature of implicit knowledge acquisition, it is desirable and legitimate to try to verify whether even less intrusive tools are capable of achieving the same results. Enhanced incidental learning is one such tool. The term was introduced by Long in his discussions on the ISLA research agenda (2017) and the optimal types of input (2020). This dissertation aims to verify its effectiveness.

The term 'enhanced incidental learning' refers to an internal process taking place in the learner's mind, whose psycholinguistic underpinnings can be related to Ellis' (2001) claims about the Law of Contiguity. According to this basic principle, when objects are experienced together, the mind tends to associate them so that when one of them is perceived, the other is evoked, as well. Applying this law to formulaic language learning, what follows is that long-term memory is statistically sensitive to formulaicity. In other words, two vocabulary items that often co-occur in the input become associated in the learner's mind. According to Ellis, this process takes place in two steps. First, a new FS is encountered in the input. At this point, a conscious association needs to be made, after which the presence of this first trace makes it possible to unconsciously detect subsequent encounters with the sequence. This engages a tallying process affected by frequency, capable of creating implicit knowledge of the association through statistical learning.

Pedagogic practices aimed at triggering enhanced incidental learning should be designed to create the best learning conditions possible for these processes to take place.

The first phase of learning requires greater consciousness in order to create an association between the formulaic sequence components. Therefore, the subject's attention should be drawn to the first one or two occurrences of the target sequence by means of input enhancement devices, such as typographical or aural enhancement. This fosters the creation of a first memory trace of the FS. Crucially, in the present work a further step is attempted. Even though Ellis claimed that conscious noticing of the first occurrence is necessary for subsequent implicit tallying to take place, here the hypothesis is promulgated that unconscious detection might also trigger the process. If so, the learning conditions would be optimal, since not involving conscious noticing would make it more likely to result in the acquisition of implicit knowledge.

In order to test this hypothesis, in the present study three different kinds of input enhancement are added to the first two occurrences of the target items. These three types of input enhancement present different degrees of noticeability, and their relative effectiveness is measured.

Crucially, even though input enhancement devices are employed, a core difference exists from their traditional use. Input enhancement is usually meant to result in noticing, that is, intentional learning, which in turn creates mainly explicit knowledge (see § 3.3). However, here, a precise and limited use of input enhancement tools is meant to foster detection, an unconscious process likely to result in implicit knowledge.

This aim is pursued by setting incidental learning conditions for the second phase of the pedagogic treatment. To boost the implicit tallying of the co-occurrences, incidental learning conditions are re-established. Namely, in the following encounters of the target sequence, the enhancement devices are removed, and statistical learning is supported by a high frequency of occurrence. Among the different options of incidental learning conditions, in the present study bimodal presentation, i.e. reading-while-listening, has been adopted.

The next sections deal in detail with the single pedagogic tools involved in enhanced incidental learning.

3.2. Constructed salience and input enhancement

As already mentioned (§1.2), salience is considered as one of the factors capable of affecting language learning, because "salient items or features are attended, are more likely to be perceived, and are more likely to enter into subsequent cognitive processing and learning." (Ellis 2018, p. 21). According to Ellis' (2001) Law of Contiguity, the first phase of formulaic language learning is the creation of a trace by means of conscious noticing of the unknown sequence. Only afterwards can implicit knowledge be developed through statistical learning. Manipulating the salience of the first occurrence of a new formulaic sequence is therefore meant to facilitate the first phase of learning. With this aim, the present study controls salience as an independent variable.

Researchers and language teachers have developed different kinds of interventions meant to increase perceptual salience. It is relevant to follow the evolution of the rationales underpinning these pedagogic techniques, since much of it led the design of the present dissertation. Sharwood-Smith (1981) was the first to envision pedagogic intervention in terms of salience and raising language consciousness in the classroom. Noticeably, Sharwood-Smith's line of argument started with assuming that "[t]he ultimate, most highly prized goal of learning, i.e. spontaneous, unreflecting language use, is uncontroversial" (p. 159). In other words, the main aim of language teaching should be the creation of implicit knowledge. However, the role of consciousness in the process was (and is) a matter of debate. He pointed out that even though purely communicative methods with no promotion of conscious awareness of the language are often advocated as the only or best way to create implicit knowledge, such practices require an amount of time and input that is rarely available to teachers. This is exactly the concern that in the present dissertation leads to the proposal of enhanced incidental learning (see infra § 3.1; Long 2017). In this regard, Sharwood-Smith highlights the potential benefits from a practice meant to raise the learner's consciousness of the formal aspects of language. Such practice can be seen as a resource the learner (especially the mature learner) can exploit to speed the process of learning how to communicate in the target language. Sharwood-Smith (1981) proposes a classification of the possible pedagogic interventions:

> Strictly speaking, the discovery of regularities in the target language whether blindly intuitive or conscious, or coming between these two extremes, will always be self-discovery. The question is to what extent that discovery is guided by the teacher. The guidance, where consciousness raising is involved, can take more or less time or space and it can be more or less direct and explicit. (pp. 160-161)

Consciousness-raising techniques can be assigned to one of four types, according to their level of elaboration and explicitness:

- Type A activities have a high level of elaboration while being less overt (low explicitness). For instance, learners are required to go through a sequence of structured stages in order to get to use a rule or pattern in the language, but they are not initially aware of the metalinguistic aim of the activities.
- Type B practices have both a high level of elaboration and explicitness; therefore, the students might go through the same steps as in Type A, but with more awareness of the grammar-related goal.
- Type C are the most implicit, with low elaboration and explicitness, mainly consisting of brief and indirect clues. The present project focuses on this kind of intervention.
- Type D are highly overt while requiring a limited level of elaboration, like in the case of providing the learners with ready-made metalinguistic prescriptions.

Sharwood-Smith (1991) later introduced the term 'input enhancement' as a safer replacement for 'consciousness raising'. The rationale for this change lay in the observation that "what is made salient by the teacher may not be perceived as salient by the learner" (p.120); therefore it is more accurate to focus the terminology on the intervention (input enhancement), rather than on the internal mental processes of the learner (consciousness raising), which should be the subject of empirical investigation.

Sharwood-Smith points out that externally created salience does not necessarily imply an effect on language development. Indeed, "salient" input may not be experienced as such, and even if noticed it may have no effect on learning. Finally, interventions may affect metalinguistic knowledge without modifying internal grammar, or vice versa. In other words, enhancing the input does not necessarily mean that input becomes intake for meaning or for acquisition (Sharwood-Smith 1993). This line of argument highlights the need to investigate the actual effect that input enhancement interventions have on learning, at both the process and the product level. This need implies the employment of tools capable of measuring both explicit and implicit knowledge, and it constitutes one of the main goals of the current project.

A wide range of interventions are included in the term 'input enhancement'. Besides the explicitness-elaboration classification, another core distinction is between negative and positive enhancement. Negative input enhancement constitutes negative evidence and aims at highlighting what is not acceptable according to standard norms. This can be achieved by somehow flagging forms as incorrect (e.g., by underlining, coloring, attaching an asterisk), and it generally requires further signals and information to be brought to the attention of the learner. Negative enhancement is often carried out on learner-produced input (with or without a garden path procedure), thus adopting a hypothesis-testing model. Therefore, it can also be described as corrective feedback.

On the other hand, positive input enhancement provides positive evidence. Positive evidence usually consists of naturally occurring samples of correct language. Emphasizing the correct form through input enhancement aims to make it more salient, so as to trigger a change in the knowledge the learner has of the structure. Such emphasis may be achieved in many different ways.

Textual aspects such as frequency can be manipulated, artificially increasing the number of occurrences of a given structure or vocabulary item (input flood). On a different level, morphemes, words and phrases can be made typographically more salient through bolding, underlining, italicizing, changing the font type or size, etc. A different kind of visual enhancement may involve adding images or pictures referring to specific items in the text. Aurally, salience can be increased by borrowing strategies naturally adopted by native speakers when dealing with language learners: adding pauses before and after the target items, increasing volume, and slowing reading pace. Finally, technology provides researchers and practitioners with new tools, such as interactive links to glosses and translations.

Of the numerous options of interventions meant to enhance the input, the current study focuses on positive evidence with low explicitness and elaboration (Type C in Sharwood-Smith's 1981 classification). This choice is motivated by the need to maintain incidental learning conditions, which in turn can result in the creation of implicit knowledge. Low explicitness and low elaboration are the features most likely to lead the speaker to detect the target items without triggering intentional and explicit learning. In view of these considerations, the enhancement tools employed in the current experiment are increased frequency, typographical enhancement, and aural enhancement. All are unobtrusive with different levels of noticeability, allowing for more detailed investigation of the effects of awareness on the learning process.

Below, each device is dealt with more deeply through a review of the available empirical findings. Existing research both provides methodological guidance and points to gaps to be further investigated.

3.2.1. Positive input enhancement with low explicitness and elaboration: empirical evidence and research gaps

Numerous studies have empirically investigated the effectiveness of Type C positive enhancement, but the existent evidence is often mixed.

3.2.1.1. Increased frequency

The first device employed in this study to manipulate constructed salience of given linguistic forms consists of artificially increasing the forms' frequency in a text. In addition to increasing salience, augmented frequency also makes it more likely for learners to trigger statistical and therefore implicit learning (for a discussion of statistical learning, see § 2.3).

Numerous studies have investigated the effect of frequency of exposure on language acquisition. On a pedagogical level, increasing the frequency of target items in the input is a relatively simple and widespread practice, although considerable time is needed to produce new versions of a text with alternate frequencies of target forms.

Therefore, it is highly desirable to provide both empirical confirmation of its effectiveness and practical indications for instructors about such details as the minimum number of occurrences necessary to engage learning and acquisition.

Research addressing the effects of input frequency on vocabulary learning have often found a positive correlation between number of exposures and learning. One of the first studies providing evidence for a positive effect of repetition on learning was the well-known experiment by Saragi, Nation, and Meister (1978). Researchers had native speakers of English read Anthony Burgess' A Clockwork Orange and then carried out a surprise vocabulary test about the Russian-based slang in the text. Results showed that more frequently occurring words were learned better.

Horst, Cobb and Meara (1998) replicated Saragi, Nation and Meister's (1978) work on A Clockwork Orange, working with L2 speakers and controlling for book length and word frequency. Their subjects engaged in reading while listening to a 109-page book with 45 unknown target words over a period of ten days. Findings showed no effects for words' general frequency, but the number of occurrences in the text did affect learning, with greater results for items that were repeated eight or more times in the text.

Rott (1999) manipulated frequency as an independent variable. L2 learners were exposed to between 24 and 36 sentences containing two, four, or six occurrences of the target words. Productive and receptive offline tests of vocabulary knowledge showed a learning effect in the two-repetition condition, with no significant difference between two and four repetitions. Conversely, subjects exposed to six occurrences of the target words significantly outperformed the two- and four-repetition groups.

Similarly, Waring and Tataki (2003) found a clear advantage for more frequently occurring words when testing L2 speakers on the new words encountered in a 400-headword graded reader. They divided the words into five frequency bands: 1, 4-5, 8-10, 13-14 and 15-18. Their results showed significantly different learning rates for words in the single occurrence and the 4-5 occurrences bands. Significantly more words repeated 8 to 10 times were learned than those occurring 1 or 4-5 times. Finally, no difference was found between words repeated 8 to 10 times and more frequent words (13-14 and 15-18 bands).

Tekmen and Daloglu (2006) repeated their 30 target words 1 to 15 times in a 2400-word text, and measured effects of word frequency and proficiency level. Word frequency in the text accounted for 29% of the variance, although some form-meaning connections were found even at the one-exposure level. However, the study procedures allowed participants to reread the text silently as many times as they wished after the reading-while-listening treatment. This undermined control of the subjects' actual number of exposures to the target words.

Webb's (2007) study provides a deeper insight into the role of frequency in vocabulary learning, as it measured six different levels of vocabulary knowledge by means of 10 different tests. English L2 speakers were randomly assigned into four experimental groups and a control group. Experimental groups were read a set number of pages, with 1, 3, 7 or 10 occurrences of the ten target nonsense words in context. The findings suggested that repetition positively affected learning, and that sizeable learning effects may happen when encountering unknown words in context at least 10 times. However, in order to gain full knowledge of a word, the number of repetitions should be higher. Chen and Truscott (2010) based their design on Webb's (2007), but emphasized the ecological validity of the study, having real words embedded in meaningful reading passages instead of Webb's controlled context. Results for repetition supported those reported by Webb (2007), while suggesting that the original study's highly controlled nature might have overestimated learning effects.

Webb and Chang (2015) carried out a long-term study investigating the effects of word frequency and distribution on the learning of one hundred target words, which were quasi-randomly selected from ten graded readers (54,000 words). Their findings showed high vocabulary gains, but no significant frequency effect. It is noteworthy that, as in Tekmen and Daloglu's (2006) study, control for incidental-learning conditions was not strict. Participants were allowed to search for unknown words in the dictionary and to read the books again during weekends. Therefore, there is no way of knowing how many times each target word was actually read by the participants. For this reason, results are more relevant for the effects of extensive reading than for reading while listening and incidental learning.

Malone (2018) employed more rigorous control over learning and exposure conditions. He embedded 32 unknown low-frequency words in four stories, with either two or four repetitions. Participants read the material in timed slides, which prevented them from re-reading the material. Learning effects were found even at the two-repetition level and frequency effects were measurable from two to four exposures. The effects of bimodal exposure were also investigated in this study, as will be discussed in section 3.3.

All of these studies provide solid empirical evidence of the positive effect repetition has on the learning of single words. However, there is clearly little agreement among different studies of the number of repetitions needed to trigger the creation of new knowledge. Indeed, significant effects were found with as few as two occurrences (Malone 2018), starting at the 4-5-repetition band (Waring & Tataki 2003), not before six occurrences (Rott 1999), with eight or more repetitions (Horst et al 1998), and only after ten occurrences (Webb 2007; Chen & Truscott 2010). Finally, results from a meta-analysis of correlational studies should be mentioned (Uchihara et al 2019). Uchihara and colleagues (2019) synthesized and quantitatively analysed 45 effect sizes from 26 studies, showing that repetition has a medium effect on incidental vocabulary learning (r = .34). In other words, they found that 11% of the variance in word-learning in incidental conditions can be explained by frequency of encounters. This indicates that frequency is important, but probably less central than often assumed in the literature, and that it is only one of many variables affecting incidental vocabulary learning.

The crucial role of frequency in formulaic language is widely recognized. Boers and Lindstromberg (2012) summarized it well, stating: "learners' uptake of formulaic sequences as a by-product of message-oriented activities alone is an incremental process that typically requires multiple encounters with the same items; it is therefore strongly contingent upon the frequency of occurrence of the items in the input" (p. 99). Despite this theoretical agreement, studies of frequency's effects on learning formulaic language are less numerous than those about single-word vocabulary items. Moreover, their findings are currently mixed with regard to the correlation between number of occurrences and incidental learning of FL; some studies have even found no effect for frequency.

Pellicer-Sánchez (2017) studied incidental learning of collocations from reading, with a focus on the role of frequency of exposure. L2 English learners read a 2300-word story containing 98% known words and either four or eight occurrences of the six target collocations. In order to address transparent collocations without a pretest, target phrases consisting of a real adjective and a pseudoword were employed. Participants' knowledge of the target items was assessed through offline tests addressing form and meaning recall and recognition. The analysis showed positive results for incidental learning, but no significant difference between the four- and eight-repetition groups. This finding may indicate that frequency has less importance than expected, or that its effects are not as linear as in the acquisition of single words. However, two points need to be taken into account when interpreting the results. First, knowledge was only assessed through offline tests, meaning that more fine-grained measurements would be necessary to assess the effects of statistical learning. Second, it is not clear how the use of pseudowords may have affected results. The use of unknown words may have diverted learners' attention from the collocation to learning the meaning of the single words.

A similar issue emerged in Szudarski and Carter's (2014) study. Participants read six stories and were then tested on 20 target collocations. Ten of the target items occurred once in each story (i.e. six total exposures), and the other 10 occurred twice in each story (i.e. 12 total exposures). Results from offline tests for form and meaning recall and recognition showed no significant difference between the six- and twelve-exposure items. In this case too, the authors speculated that the result might be related to the fact that all of the target collocations contained infrequent words unlikely to be known to participants. Therefore, learners may have faced a double-learning task, i.e. they needed to learn both the single words and the collocations. This limitation is accounted for and overcome in the present study (§ 4.3).

Webb, Newton and Chang (2013) carried out the experiment that serves as the main methodological reference point for the present dissertation. One hundred and sixty-one students learning English as a foreign language were exposed to reading-while-listening to a modified version of a 700-headword graded reader. As the focus of the experiment was the role of frequency, four different versions of the instructional material were created, containing 1, 5, 10 or 15 occurrences of each of the 18 target collocations. Crucially, the collocations were relatively opaque and made up of known words, which avoided the double-learning-task issue of the aforementioned studies (Pellicer-Sánchez 2017, Szudarsi & Carter 2014). The sample was randomly divided into four experimental groups, each exposed to one version of the instructional material, and a control group which only performed the tests. The authors tested prior knowledge of the opaque target collocation with a form-recognition pretest, which was administered one week before the treatment. The posttests included the same test format as the pretest and three more offline tests, assessing productive knowledge of form and receptive and productive knowledge of meaning. Results showed that frequency had a significant effect on learning, with knowledge increasing as the number of repetitions increased. Sizable effects on learning were found for 15 encounters.

Webb and colleagues' findings provide support for the role of statistical learning of L2 collocations. Namely, it adds to the evidence in favor of positions such as Durrant and Schmitt's (2010), which holds that adult L2 learners retain the ability to engage in statistical, implicit learning (§ 2.3). As a consequence, further support is provided for the applicability of Ellis' Law of Contiguity not only to L1 but also to L2 FL acquisition. However, overall the existing literature suggests that the learning of formulaic sequences is a more complex process and involves different factors from those affecting single words. Keeping this in mind, the present study addresses the methodological limitations of the previous experiments, employing formulaic sequences composed of known words while avoiding a pretest (§ 4.3.2; 4.5).

It needs to be pointed out that it is possible for frequency not to have its anticipated role, especially when it comes to idioms. Indeed, idioms' features are believed to foster acquisition after a brief exposure. According to Reuterskiöld and Van Lancker Sidtis' (2013) results with L1 children, the unitary, non-compositional and linguistic-contextual characteristics of idioms can lead to acquisition from a single exposure. Given the mixed nature of findings about the role of frequency for the learning of collocations and idioms, more data investigating the effects of frequency with idioms are needed. The present study investigates this issue by monitoring eye-movements of a subsample of subjects at the process level, measuring how familiarity with idioms changes from the first to the seventh encounter.

3.2.1.2. Aural enhancement

The second tool under investigation here is aural enhancement. This involves the unobtrusive manipulation of listening materials with the aim of making specific linguistic forms more salient. As mentioned above, it can include increased volume, slower pace, or short pauses added before and/or after the target items. Aural enhancement has great potential for the aims of the present research for at least two reasons. First, it is ecologically valid, as this kind of input manipulation (namely, the added pauses) has been found to occur naturally occur in native-to-nonnative communication (Long 1983). Second, aural enhancement is less obtrusive than typographical enhancement or unnaturally high frequency of target items. Therefore, it is more likely to trigger unconscious detection rather than intentional learning and thus to result in implicit learning and knowledge. Despite these promising features, empirical studies on aural enhancement's effectiveness for learning are rare.

Cho and Reinders (2013) published the most relevant study of the effects of aural enhancement on learning of a grammatical target item (passive form). Seventy-two L2 learners of English from three classes were given the audio version of a graded reader (a 90-minute recording) to listen to autonomously over one week. The three intact classes were randomly assigned to one of three conditions: the pause group (1.5 seconds pause digitally inserted before and after the target form), the reduced-speed group (target recording slowed down by 10%), and the control group (no enhancement). Learning was measured by means of a timed grammaticality judgment test, which is considered to have the potential of tapping into implicit knowledge. While all three groups showed improvement from pretest to posttest, no significant differences were detected between the experimental groups and the control group, nor between the two kinds of aural enhancement. The authors discuss this outcome, pointing out factors such as the limited amount of input, the short time period over which it was presented, the complexity of the target form and the difficulty of processing aural-only input. However, there are other possible limitations in Cho and Reinders' (2013) study, which the current study was designed to overcome. First, as students were required to listen to the text at home, no control over exposure to the input was provided. Therefore, in the current study, the subjects were only exposed to the input during the experimental sessions. Second, a 1.5-second pause is highly unnatural and thus undermines ecological validity, one of the most desirable features of this treatment. In the present experiment, the pause is reduced to a more natural 0.4-second duration, a closer approximation to the 'one-beat pause' observed in foreigner talk discourse (Long 1983). Finally, agreeing with the authors about the possible difficulties of an aural-only presentation of the input, subjects were provided with the written form, as well.

Other studies have touched on the topic of aural enhancement, but their methodological choices make them less relevant to the aims of the present investigation. For instance, Negari and colleagues (2018) addressed the effectiveness of aural enhancement on EFL learners' retention of intensifiers. They found a significant effect for aural enhancement on learning; however, the treatment received by their control group differed from the experimental group not only in lack of exposure to aural enhancement, but also in lack of any aural input at all. In fact, the experimental group was exposed to reading-while-listening plus aural enhancement (increased volume), while the control group only read the instructional text with no aural component. This makes it impossible to distinguish the effects determined by aural enhancement from those related to reading while listening.

A similar flaw in the design makes Zanjan's (2017) findings only partially relevant. That study compared the effectiveness of aural (emphatic stress) and textual (italics and boldface) enhancement on explicit grammar knowledge gains. Unsurprisingly, in the offline posttest, textual enhancement resulted in better scores than aural enhancement, which suggests that the more noticeable nature of textual enhancement is more likely to affect explicit knowledge. However, the design did not include a control group, so even though descriptive statistics show improvements from pretest to posttest for subjects exposed to aural enhancement, there is no way of knowing whether such gains are due to the experimental treatment, to mere exposure to the target items, or to a testing effect.

No study appears to have investigated the effects of aural enhancement on the learning and acquisition of formulaic sequences. The present research aims to address this gap.

3.2.1.3. Typographical enhancement

The last kind of enhancement examined here is typographical enhancement, about which more empirical data are available. Its simplicity of employment makes typographical enhancement very common in classroom materials, where it is often the starting point for a more explicit elaboration of language forms. However, provided it is not followed by teacher-initiated focus on forms, typographical enhancement has an unobtrusive nature that could potentially result in incidental learning while speeding up detection.

Most empirical studies addressing typographical enhancement have focused on the learning of grammar. Findings do not always confirm the advantage of exposing learners to input enhancement.

A study by Doughty (1991) was one of the first experiments to investigate the effectiveness of salient visual clues. With a between-group, pretest-posttest design, she examined acquisition of English relativization by L2 learners randomly assigned to three groups: a meaning-oriented instructional group, a rule-oriented instructional group, and a control group. The three groups read the same texts containing the target forms, with the instructional tools added to the text placed at different points on an explicit-implicit continuum. The control group was exposed to the most implicit treatment, since salience of the target structure was only due to redundancy (high frequency) and markedness of the relative clauses. In the meaning-oriented treatment, further salience was added to the target structures as they were highlighted and capitalized in the text, and lexical information was available. Finally, the rule-oriented group received the most explicit treatment, as in addition to the salient visual clues, it also included metalinguistic descriptions. The scores of both written and aural tests showed a clear advantage for the treatment groups over the control group. Crucially, no significant difference emerged in the grammar knowledge gains between the subjects exposed to salient visual clues only and those who also received metalinguistic instruction. Moreover, only the meaning-oriented group outperformed the control group in a test of the text meaning, while the rule-oriented group showed poor comprehension of content. Such findings allowed the researcher to hypothesize that perceptual salience may have been the main factor in the success of instruction, while the explicit grammar explanation was not.

Findings from this early study are only partially confirmed in the following empirical literature. Indeed, while some empirical studies found grammar learning effects associated with textual enhancement (e.g., Cintrón-Valentín & Ellis 2015; Jahan & Kormos 2015; Issa et al 2015; LaBrozzi 2016; Lee 2007), other experiments did not (e.g., Indrarathne & Kormos 2016; Izumi 2002; Loewen & Inceoglu 2016; Winke 2013). A meta-analysis by Lee and Huang (2008) found only a negligible effect (d=0.22) for typographical enhancement in grammar learning. Even though this outcome is often quoted as a point against input enhancement (e.g., Leow & Martin 2018), it is crucial to point out that the effect size was calculated by contrasting experimental groups with input flood groups, rather than with actual control groups. Such an effect size therefore does not communicate the effectiveness of input enhancement as such, but rather the difference between two kinds of input enhancement: typographical enhancement and artificially increased frequency.

Despite the mixed results on L2 development, empirical findings show more agreement when it comes to the effects of typographical enhancement on attention allocation. Three main tools have been employed in order to investigate how typographical enhancement affects attention: notetaking (Izumi 2002), think-aloud protocols (Bowles 2003; Leow 2001; Leow et al 2003) and eye tracking (Cintron-Valentin & Ellis 2015; Indrarathne & Kormos 2016; Issa et al 2015; Loewen & Inceoglu 2016; Simard & Foucambert 2013; Winke 2013). The majority of these studies confirmed that typographical enhancement increased the amount of attention paid to the target items.

Izumi (2002) asked his subjects to take notes while reading the input text, which was either enhanced or unenhanced according to the experimental group participants were randomly assigned to. The analysis of the notes showed a significant effect for input enhancement in the augmented noticing of the target structure, i.e. relative clauses. Think-aloud protocols applied to textual enhancement investigation present mixed findings, with studies reporting benefits (e.g., Bowles 2003) or no effects (Leow 2001; 2003). However, both notetaking and think-aloud protocols have limitations as online means to assess the level of awareness. Indeed, this kind of practice is based on verbalization, which is assumed to be closely associated with awareness. However, this assumption is not necessarily warranted, since a lack of verbalization does not provide strong evidence for unconsciousness, and "awareness may happen more quickly than concurrent verbalization allows expression of" (Rebuschat et al 2015, p. 303). At the same time, the presence of verbalization cannot rule out the possibility that some unconscious process is concurrently taking place. Moreover, the very request to produce a verbalization is extremely likely to affect and modify the process.

For these reasons, it is even more important for findings on input enhancement and its role in awareness and attention to be confirmed in online studies employing a more fine-grained process measure, i.e. eye tracking. Even if product-level tests fail to highlight new knowledge, eye movements measured at the process level in different studies show the effectiveness of textual enhancement in terms of the amount of attention paid to target items.

Winke (2013) employed eye tracking to investigate the effects of typographical enhancement (red coloring and underlining) on reading behavior, content understanding and grammar learning. When comparing outcomes from enhanced and unenhanced texts, the analysis of fixations and regressions showed that enhancement augmented both, which implies increased attention allocated to the passive target forms.

Similar results are reported in Simard and Foucambert's (2013) study, which addressed both online (eye tracking) and offline (verbal reports) measures of noticing. Eye-tracking measures showed increased consciousness in participants when reading enhanced compared to unenhanced input. Interestingly, no correlation was found between online and offline assessments of noticing, further evidence of the two tools tapping into different dimensions of awareness. As a consequence, it is clearly desirable to include both eye-tracking and retrospective verbal reports in empirical studies, and then to triangulate the findings (Rebuschat et al 2015). The present experiment follows this recommendation.

Further support for the effect of typographical enhancement on awareness comes from studies that compared it not only to an unenhanced control condition, but also to other kinds of instruction. Indrarathne and Kormos (2016) randomly assigned their participants to one of four experimental groups, which were exposed to (i) typographical enhancement (boldface), explicit instructions to pay attention to the enhanced words and metalinguistic explanation; (ii) typographical enhancement and explicit instructions; (iii) typographical enhancement with no explicit instructions; or (iv) unenhanced text. Eye-tracking showed that enhancement, even without explicit instruction for participants to pay attention to it, resulted in longer fixations on the target items (causative had constructions), i.e. it positively affected levels of consciousness.

Issa and colleagues (2015) compared the effects of typographical enhancement (red coloring) and structured input activities on the reading and learning of Spanish direct object pronouns. The rationale was that while input enhancement is an external manipulation of attention, structured input activities operate at an internal-salience level (e.g., VanPatten 2004). Analysis of eye-tracking data on skipping rates showed that both interventions significantly improved the amount of attention paid to the target items as compared to the control group, with no difference between input enhancement and structured input activities.

Cintrón-Valentín and Ellis (2015) investigated different kinds of focus on form, potentially capable of assisting the learners in overcoming blocking and learned attention (see also § 1.2). Subjects were randomly assigned to three experimental groups and instructed about Latin verb morphology in one of three ways: explicit grammar instruction, typographical enhancement (verb inflections highlighted in red and bold), or verb pretraining with an additional introduction involving English translations. Eye tracking data demonstrated that typographical enhancement led to significantly more scrutiny of the verbs compared to a control group, with no differences from the more explicit treatments involving grammar instruction and pretraining.

To summarize, a large number of empirical studies have investigated the effectiveness of typographical enhancement on grammar learning. However, a review of this literature shows mixed findings.

The apparently limited effectiveness of typographical enhancement on the creation of new grammatical knowledge has been explained in the literature in terms of a difference between quantity and quality of attention (Izumi 2002; Leow & Martin 2018). Indeed, tools such as eye tracking, think-aloud protocols and notetaking demonstrated an increase in noticing and in the amount of attention paid to the target items. However, such attention sometimes failed to result in the creation of significant knowledge (e.g., Izumi 2002; Winke 2013). These findings are interpreted as evidence of typographical enhancement triggering a superficial kind of attention, resulting only in sensory registration with no persistence of memory traces. From a slightly different perspective, it is claimed that typographical enhancement is likely to prompt only semantic processing, which is not deep enough for the internalization of grammatical information.

Moreover, the data on awareness come from the beginning of the learning process, which may not yet be detectable in offline, coarse-grained immediate posttests. Indeed, very few studies employed long treatments and delayed posttests, although integrating new grammatical information requires numerous occurrences and time.

While such considerations are valid with regard to grammar learning, they are not necessarily generalizable when it comes to formulaic language. Empirical studies of typographical enhancement and FL learning, albeit less numerous, present a higher degree of agreement as to the significant and positive impact of enhancing formulaic sequences in written input.

Peters (2012) investigated the effects of instruction and typographical enhancement (boldface and underlining) on the learning of formulaic sequences and single words embedded in a 1148-word reading text. The study had both between-subject and within-subject variables. The between-subject variable was type of instruction; subjects were randomly assigned to two experimental groups, only one of which received explicit instruction to pay attention to new formulaic sequences and single words. Enhancement was the within-subject variable, as only half of the 12 target formulaic sequences and half of the 12 target single words were enhanced. The dependent variables were subjects' scores on pretest and immediate and delayed posttests, consisting of form and meaning recall (L1 to L2 translation) of the target items. Peters found that enhancing the target items significantly improved learning, while explicit instruction did not. Moreover, the analysis of the interaction between the variables showed typographical salience to be especially effective for the learning of formulaic sequences, as compared to single words. According to the author, this might be related to the learners' tendency not to recognize formulaic sequences as wholes when they are found in the input, especially if they are semantically transparent and composed of known words (Nation 2001). Typographical enhancement directly affects this aspect, preventing students from overlooking the formulaic sequences and improving the amount of attentional resources allocated.

These findings are confirmed in the above-mentioned study by Szdudarki and Carter (2016), who compared two kinds of instruction: input flood only, and input flood plus typographical enhancement (underlining). Following a between-subject, pretest-posttest design, 51 subjects were randomly assigned to one of the two experimental groups or to the control group. Over three weeks, participants in the experimental groups read for content the same six stories, in which twenty target collocations were embedded either six or twelve times. The only difference between the treatments was the presence or absence of typographical enhancement. Assessment took place two weeks before and two weeks after the treatment and included five offline tests of productive and receptive knowledge of form and meaning. Analysis of the scores suggested that only typographical enhancement resulted in the creation of new knowledge, while the input-flood-only treatment did not, for both the six- and twelve-occurrence frequency bands. Not surprisingly, receptive tests showed better results than productive tests. Such findings confirm the effectiveness of typographical enhancement for the learning of formulaic sequences, while pointing out a methodological issue. As mentioned earlier, the authors hypothesize that the apparent inefficacy of purely incidental conditions (input flood) and the lack of productive knowledge might be due to the target collocations being composed of infrequent words, probably unknown to the participants. This imposed a double learning task on the subjects, as they had to face both new words and new collocations, which might have reduced the frequency effects. This limitation is taken in consideration and overcome in the present study, which employs unknown formulaic sequences composed of known words as target items (§ 4.3.1 and 4.3.2).

Boers and colleagues (2017) had a slightly different focus, as they did not include input flood in their investigation but rather focused on the potential ability of typographical enhancement to foster learners' sensitivity to formulaic sequences in general, i.e. extending its benefits to unenhanced sequences. In order to do so, they randomly assigned 81 subjects into three experimental groups, all of which read the same two texts for content, each composed of 400 words and containing one occurrence of the 16 target formulaic sequences. Three versions of the text were created for the three experimental conditions: (i) all 16 target items enhanced (underlining); (ii) only 8 out of 16 target items enhanced; and (iii) no enhancement. Learning was assessed through an episodic memory test. Results showed that underlining was effective in improving the learning of formulaic language. However, such benefits did not extend to the unenhanced items, which suggests that typographical enhancement did not boost sensitivity to the formulaic dimension of the text beyond the enhanced items. In fact, subjects exposed to only eight enhanced items scored worse on the unenhanced formulaic sequences than the no-enhancement group. This may point to an undesirable side-effect of input enhancement called trade-off, which has also been observed in the empirical literature on grammar forms (e.g., Lee 2007; Overstreet 1998). When a trade-off effect occurs, the attentional resources allocated to enhanced items are subtracted from other aspects of the text, thus hampering learning of unenhanced forms or comprehension in general. In Boers and colleagues' study, however, these differences fell short of statistical significance, so more experimental data are required to claim the existence of a trade-off effect related to typographical enhancement of formulaic sequences.

Especially relevant to the present study are experiments which employed not only offline but also online measurements, aiming to assess gains in both explicit and implicit knowledge. Sonbul and Schmitt (2013) employed a counterbalanced, within-subject design to compare three learning conditions: enriched (input flood, i.e. three occurrences in a short passage), enhanced (i.e. same as enriched condition, with bold and red font added), and decontextualized (i.e. collocations presented individually on slides to be memorized). Forty-two subjects carried out two offline test to assess receptive and productive explicit knowledge, while implicit knowledge was measured by means of priming. In this online test participants are presented with the first word of the collocation as the prime and the second as the target, and they have to decide whether the second string of letters is a real English word or not. The test battery was repeated immediately after the treatment and then again two weeks later. Results of the offline tests showed that all three experimental conditions led to significant and durable learning at both productive and receptive levels. The enhanced condition was more effective than the enriched condition, and as beneficial as the decontextualized condition. Therefore, this study is in line with the literature in claiming the benefits of typographical enhancement for explicit knowledge of formulaic sequences. When it comes to implicit knowledge, however, both immediate and delayed priming sessions resulted in no measurable gains for any of the instruction conditions. This outcome might be related to the short duration of the treatment (only one session) and the number of occurrences of the target items (three), which may not be enough to trigger measurable statistical, implicit learning.

Toomer and Elgort (2019) ran a conceptual replication and extension of Sonbul and Schmitt's (2013) study. They applied the same within-subject counterbalanced design to a larger sample, as they exposed their 62 participants to three learning conditions: reading only (no typographical enhancement), bolding, and bolding-plus-glossing. Compared to Sonbul and Schmitt (2013), the treatment was longer and spaced (three reading sessions over two days), and target collocations occurred a total of nine times. These choices were meant to boost statistical learning and implicit knowledge gains. As in the previous study, testing sessions took place immediately after the treatment and then again two weeks later, and included both offline and online measures. Explicit knowledge was measured by offline recall and recognition tests, while lexical-primed decision was employed to detect the creation of implicit knowledge. Analysis confirmed Sonbul and Schmitt's outcomes: the experimental groups developed significant explicit knowledge of the target collocations, with the enhanced condition more effective than the reading-only condition. Findings on implicit knowledge extend and partially confirm those of the previous study. No priming effect was shown as a result of typographical enhancement, while implicit knowledge emerged for collocations in the reading-only condition. The fact that learners were able to develop implicit knowledge in this study and not in Sonbul and Schmitt's (2013) can be explained by the longer, spaced nature of the treatment, as well as the increased frequency of the target items. The lack of implicit knowledge as a result of the enhanced condition, however, raises a more complex issue that is highly relevant for the present research. The authors hypothesize that as typographical enhancement made the target collocations more salient (as the explicit knowledge gains demonstrate), this might have interfered with the process of word-to-text integration (Perfetti et al. 2008). In other words, implicit knowledge of formulaic language is created by means of statistical learning, i.e. unconscious detection of co-occurrences of word sequences. Bolding is likely to force the allocation of additional attention to the enhanced items, thus changing the nature of the process and hampering its implicit, tacit and unconscious nature.

Empirical evidence of such an attentional shift comes from Choi's (2017) eye tracking study. Thirty-eight English L2 speakers were randomly assigned into two groups, which read for content one of two versions of the same text. In the first version, the 14 target collocations were enhanced (bold typeface); in the second they were not. The researcher recorded subjects' eye movements, with a region of interest of the whole collocation and a focus only on late measures, i.e. total reading time and total fixation count. Collocational knowledge was tested before and after the treatment with offline recall tests. The analysis of fixations and pretest and posttest scores showed that typographical enhancement was effective in increasing the amount of attention paid to unknown collocations and resulted in better knowledge gains. In greater detail, eye-tracking data revealed that that participants in the two groups did not differ in reading behavior of known collocations, while the enhancement group spent significantly more time on unknown sequences than the baseline group. Additionally, subjects exposed to bolded collocations outperformed their colleagues who read the plain test on the posttests. However, the overall results also highlight a significant trade-off effect, as the enhancement group recalled less unenhanced text than the baseline group. On the one hand, this further confirms the effectiveness of typographical enhancement in allocating attention to target items; on the other, it is a potentially problematic issue in terms of general content comprehension. Unfortunately, Choi's (2017) study did not include a measure of implicit knowledge gains, which would have shown whether this trade-off also affected the unconscious, statistical parsing processes essential for implicit learning.

Summing up, the existing empirical data confirm the effectiveness of typographical enhancement on explicit learning of formulaic sequences. When it comes to the creation of implicit knowledge, however, the sparse literature (two studies only) shows no effects (Sonbul & Schmitt 2013) or even a detrimental effect (Toomer & Elgort 2019) of typographical enhancement. Given that implicit knowledge is recognized as the most desirable goal for language instruction, due to its online availability, this point needs further experimental investigation. Moreover, only one study (Choi 2017) addressed the way typographical enhancement affects online reading behavior and attention allocation, so the positive results reported need confirmation. Finally, no study to date has employed both online measures at the process level (eye tracking) and assessment of implicit knowledge gains in relation to input enhancement and formulaic language learning. This crucial gap is addressed in the present research.

The review of the available empirical evidence concerning the three enhancement devices under consideration results in a complex picture. Studies of artificially increased frequency report mixed findings, with no study employing measures for implicit knowledge gains. Aural enhancement results in little or no effect, but the experimental data available are extremely limited. Typographical enhancement, on the other hand, is apparently beneficial for the creation of explicit but not implicit knowledge. Such mixed findings reflect the nature of input enhancement, which can be seen as a somehow contradictory pedagogical tool. On the one hand, its goal is to draw the learners' attention to specific forms in the input; on the other, it aims to do so unobtrusively and implicitly. In other words, the purpose is to increase the level of consciousness but still keep it under the awareness threshold, i.e. at the level of unconscious detection, not conscious noticing. Such a subtle task needs fine-grained assessment tools and specific experimental designs, which are not always available in the existing literature. This might help explain the mixed picture.

The contrasting nature of input enhancement is discussed in Pellicer-Sánchez and Boers' (2019) review of the pedagogical approaches to formulaic language teaching. Although they organized their chapter according to the traditional incidental/intentional dichotomy, they had to acknowledge that input enhancement could not fit in either category. They observed that numerous studies involving attention-drawing techniques such as input enhancement still expect the subject to focus mainly on a text's meaning and content. This condition does not qualify as intentional, as no explicit instruction or request to learn the enhanced FSs is provided. At the same time, enhancement is meant to move the subjects' attention from content to form, fostering noticing (Schmidt 2010), and therefore the 'incidental' category cannot apply, either. Therefore, they created a third category specifically for input enhancement, *semi-incidental* learning conditions.

Crucially, this apparent contradiction turns into a potentially powerful tool for language instruction if the instructor pursues enhanced incidental learning. Enhanced incidental learning combines some of the tools designed for input enhancement (e.g., increased frequency, aural or typographical enhancement) with incidental learning conditions, with the aim of unconscious detection rather than conscious noticing. The specific kind of incidental learning condition adopted here are the object of the next section.

3.3. Incidental learning conditions: Reading while listening

Reading a text for content is a typical example of an incidental condition for language learning. While the reader is focused on the text's meaning, unknown vocabulary items can be detected or noticed in context and learned. This kind of treatment can be beneficial on many levels. As no explicit language instruction is provided, the process is likely to be incidental and therefore implicit knowledge may be gained. This possibility is boosted if the text is long enough to trigger statistical learning through repeated encounters with the unknown items. Moreover, as Pellicer-Sánchez and Boers (2019, p.166) point out, "incidental conditions, such as extensive reading, provide additional opportunities for skills development (such as reading comprehension and reading fluency)" if compared to intentional learning, which leads the reader to focus on language forms in a more isolated fashion. Finally, encountering the same words or phrases in different contexts results in deep and transferrable knowledge.

Starting with Saragi, Nation and Meister's (1978) experiment with A Clockwork Orange (see §3.2.1), numerous studies have investigated the effectiveness of incidental vocabulary learning through natural reading (e.g., Pitts et al 1989; Day et al 1991; Hulstijn 1992). Results suggested that learning did occur, but at low rates. In order to achieve a more effective pedagogical treatment without losing the incidental condition-related benefits, factors affecting learning can be manipulated. One of these factors is exposure modality, and this explains the rationale for employing bimodal exposure, e.g. reading while listening.

In reading while listening, learners read a text while being simultaneously exposed to the aural version of the same text, which can be either recorded or read aloud by the teacher. Receiving both written and aural input has several benefits when it comes to incidental learning. First, the aural component provides a superimposed pace, which discourages the learners from stopping and intentionally learning unknown words or phrases. In Horst and colleagues' (1998) words, "reading aloud created the circumstances for incidental acquisition by precluding opportunities for intentional word learning" (p. 211).

Second, bimodal exposure can provide the learners with implicit information about chunk segmentation. Indeed, the person reading aloud naturally pronounces phrase components together, i.e. without pauses or hesitation between them (Bybee 2002), and this clarifies the formulaic nature of sequences. Therefore, information about the collocational properties of words is made available and the learner is encouraged to overcome word-for-word strategies. In other words, prosody aids the parsing of texts, by signaling the boundaries of semantic units, including formulaic sequences (Lin 2012). Since this kind of data is not available in written-only input, bimodal exposure may be especially effective for formulaic language detection and learning.

Another advantage of reading while listening concerns content comprehension. The availability of prosody and sound-symbol correspondence provides learners with an additional tool for text comprehension (Tekmen & Daloglu 2006). In turn, a better comprehension of the context results in increased cues and attentional resources for inference of unknown items' meaning. As Malone (2018) pointed out, listening while reading allows the learners to make the cognitive resources needed for phonological decoding available for form and meaning processing instead. In other words, "at a theoretical level, a facilitative role for redundant simultaneous signals could exist" (Malone 2018, p.9). A behavioral confirmation for the psycholinguistic effect of bimodal exposure comes from studies that investigated learners' preferences among exposure modality (e.g., Brown et al 2008; Chang 2009, see infra). Indeed, students claimed to be more comfortable in the reading-while-listening mode as compared to reading only and listening only.

Finally, reading while listening also presents crucial methodological benefits on an experimental design level. Having experimental subjects both read and listen to a text ensures they are all exposed to the whole text, for the same amount of time. Such close control of experimental conditions is not possible in reading only. In addition, audio books are often available and widely employed in language classes, which provides bimodal exposure with strong ecological validity.

A growing number of experiments about incidental learning of vocabulary and collocations have employed reading while listening as an incidental learning condition, and their findings have confirmed its effectiveness. As mentioned (§ 3.2.1), Horst, Cobb and Meara (1998) were among the first to investigate second language vocabulary acquisition through reading. They exposed 34 low-intermediate learners to extensive reading while listening, measuring the amount of learning and its interactions with word frequency and the subjects' vocabulary size. Superimposing a reading pace through the aural component and collecting the text after each session ensured the incidental nature of the condition, which was confirmed by follow-up interviews. According to the offline posttest results, new vocabulary was learned. However, learning rates were quite low; an average of only five new words were learned in

109 pages and six one-hour sessions. Data showed that subjects' vocabulary size affected the outcome, and this is probably one key to interpreting the findings. Indeed, the authors admitted not to have checked whether the non-target words in the text were known to the subjects. This likely hampered the target words' meaning inference and learning, since the literature has demonstrated that in order for context learning to take place, at least 95% of the words need to be known (e.g., Laufer 1992; Nation 2006). This limitation is taken into account and overcome in the present study.

Tekmen and Daloglu (2006) addressed the same points, as they investigated the effects of frequency and proficiency on incidental vocabulary learning from reading while listening. Despite a shorter duration of the treatment (only one 50-minute session), results from three different proficiency-band subjects confirmed the effectiveness of reading while listening for the development of new vocabulary knowledge. Moreover, word frequency and vocabulary size had a facilitating effect.

Webb and Chang (2015) focused on the effects of long-term extensive reading on vocabulary learning, addressing the roles of frequency and distribution. Sixty-one participants read and listened to ten graded readers over a period of thirteen weeks, with pretest, posttest and delayed posttest analyses showing high vocabulary gains. Frequency and distribution of occurrence were shown not to have a significant role. The positive results about reading while listening are encouraging, although the learning conditions in this study were only partially incidental. Students were asked to re-read the texts at home after the treatment, and vocabulary search on unknown words was encouraged. The rationale for such choices was to maintain ecological validity, as these practices are likely to be found in language courses. However, results from this study are not generalizable for the investigation of incidental learning conditions.

It seems that only one study employed bimodal exposure for incidental learning of formulaic sequences. Webb and colleagues (2013, see also § 3.2.1) exposed 161 experimental subjects to reading while listening to different versions of a graded reader with 1, 5, 10 or 15 encounters with the target collocations. A battery of four offline tests showed significant knowledge gains. Frequency effect have been previously discussed (§ 3.2.1); what is relevant here is that bimodal exposure proved effective for incidental learning of formulaic language.

Crucial evidence for the effectiveness of bimodal presentation is provided by a series of studies which had exposure modality as an independent variable, i.e. they compared reading while listening with reading only or listening only. Brown and colleagues (2008) randomly assigned 35 subjects to three experimental groups, exposed to either reading-only, listening-only or reading while listening to three graded readers, for a total of three 90-minute sessions over two weeks. Outcomes from productive and receptive vocabulary tests showed that the reading-only mode and the reading-while-listening mode resulted in similar learning rates, and that both outperformed the listening-only mode. Moreover, debriefing interviews carried out after the treatment demonstrated that students were most comfortable in the reading-while-listening mode. However, it has to be noted that participants were fully informed about the vocabulary-learning aim of the study, which makes reference to these outcomes in terms of incidental learning very questionable.

In contrast, Malone (2018) rigorously designed experimental procedures meant to ensure incidental conditions, when addressing learning of vocabulary through reading-only and reading-while-listening. Eighty participants read four graded readers on timed slides on a computer screen, with or without the aural component. Form-recognition and form-meaning connection tests showed that reading while listening was significantly more beneficial for incidental vocabulary learning. Teng (2018) confirmed these results in a study assessing four dimensions of vocabulary knowledge: form recognition, grammar recognition, meaning recall and collocation recognition. Outcomes from 60 experimental subjects showed that reading while listening resulted in larger knowledge gains than reading only for all four dimensions investigated.

Webb and Chang (2012) focused on the effect of repeated reading, as their 82 subjects read or read and listened to a short text several times over two seven-week periods. Dependent measures were provided by modified vocabulary knowledge scales, which were applied to a pretest-posttest design. Outcomes showed bimodal exposure to be significantly more effective than reading only for vocabulary knowledge gains.

Webb and Chang (2020) also designed the only study comparing the effects of reading-only, reading-while-listening and listening-only on incidental learning of formulaic sequences. Their 112 experimental subjects were exposed to a graded reader during six sessions over three weeks. Seventeen target collocations were embedded in the text with different frequencies of occurrence. The sample was randomly divided into three experimental groups (reading only, listening only and reading while listening) and a control group. Results from offline immediate and delayed posttests demonstrated an advantage for bimodal exposure over both reading-only and listening-only modalities.

As mentioned earlier, meaning inference and learning from context are only possible if the learner understands the majority of the surrounding text (Hu & Nation 2000). It has been claimed that one of the benefits of bimodal exposure is improving comprehension, for instance by means of prosody (e.g., Tekmen & Daloglu 2006). Empirical data from studies addressing comprehension in different input modalities corroborate this point.

Chang (2009) compared reading-while-listening and listening-only modalities. Eighty-four subjects from two intact classes either listened to or read and listened to two graded readers (about 3000 words in total). The bimodal-exposure group outperformed the listening-only group in terms of text comprehension. In addition, as in Brown and colleagues' (2008) study, the subjects showed a clear preference for the reading-while-listening mode, as they perceived that it made the listening task easier, the duration seem shorter, and the stories more interesting.

Chang and Millet (2015) confirmed this comprehension benefit with an extensive-reading study. Sixty-four EFL students were exposed to 20 graded readers over a period of 26 weeks, either in reading-only or in reading-while-listening mode. Subjects exposed to bimodal presentation of the input showed higher comprehension rates in both the immediate and three-month-delayed posttests.

Pellicer-Sánchez and colleagues (2018) addressed a similar point, measuring comprehension in reading-only and reading-while-listening modes. New to this study, in addition to the usual offline posttests, they also employed online assessment, i.e. eye tracking at the process level. In this case, the auditory input did not result in a measurable benefit on an offline comprehension posttest. However, the eye-tracking data provide an interesting insight with regard to bimodal exposure's potentialities as an incidental learning condition. The analysis of the reading behavior (fixation duration) showed that experimental subjects spent significantly more time reading the text in the reading-only condition than in the reading-while-listening condition. This outcome can be interpreted as a fine-grained, empirical confirmation of the effect of a superimposed reading pace. Previous studies had hypothesized that auditory input could prevent learners from stopping on unknown items to learn them intentionally, i.e. it is likely to keep learning incidental (e.g., Horst et al 1998). The fact that eye-tracking data detected the gaze dwells longer on words in silent individual reading than in bimodal exposure might be seen as an experimental evidence for this hypothesis.

More eye-tracking evidence is provided by a recent study by Conklin and colleagues (2020), who addressed reading behavior of both L1 and L2 speakers in reading-only and reading-while-listening modes. In a counterbalanced design, 31 ESL learners were presented with two reading texts, with and without aural input. Outcomes did not confirm Pellicer-Sánchez et al (2018)'s findings, as the only difference in the two modes regarded regressions, which were fewer in the reading-only mode. However, a relevant finding is reported, relating the reading patterns to vocabulary size. The researchers aimed at verifying the alignment between gaze and audio, i.e. whether subjects fixated on a word at the same time they heard it. Statistical analysis showed that this alignment was scarce in general, but significantly better for subjects with a lower vocabulary size. This finding can confirm the notion that audio is an important resource for L2 readers experiencing comprehension problems.

In general, the outcomes of experimental studies of bimodal exposure empirically demonstrate the benefits expected from this modality. Various studies have employed reading while listening and reported positive results in terms of incidental vocabulary learning (Horst et al 1998; Tekmen & Daloglu 2006; Webb & Chang 2015). These findings are confirmed by studies with modality of exposure manipulated as an independent variable (Brown et al 2008; Malone 2018; Webb & Chang 2012): better learning gains are reported in reading while listening than in reading only and/or listening only. Claims about the contribution of the aural component to comprehension and to the students' comfort are also empirically confirmed (Chang 2009; Chang & Millet 2015). At the same time, the bimodality's capacity to ensure incidental learning conditions seems to find confirmation from one eye-tracking study (Pellicer-Sánchez et al 2018).

When it comes to formulaic language, few studies are reported in the literature (Webb et al 2013; Webb & Chang 2020), but their outcomes confirm the positive results for experiments addressing vocabulary.

3.4. Summary

This chapter concerned enhanced incidental learning, which is the focus of the present dissertation. Introduced by Long (2017; 2020), enhanced incidental learning is designed to boost the acquisition of implicit knowledge by improving and directing incidental learning.

This is achieved through the manipulation of the learning conditions, with the aim of affecting the cognitive processes involved. Incidental learning conditions are thus created, and improved by adding unobtrusive enhancement devices intended to boost unconscious detection. The devices employed present growing levels of noticeability, which allow to compare the effects of obtrusive and unobtrusive enhancement on detection, noticing and eventually implicit and explicit knowledge.

The rationale and empirical literature supporting the choices regarding enhancement devices and incidental learning conditions have been exposed in this chapter; the details of how these choices are implemented in the current experimental design are the object of the next.

Chapter Four: Methodology

4.1. Research questions and design overview

The gaps showed in the analysis of the existing empirical literature are addressed through an experimental design which combines original elements and established procedures.

The main aims of the study are on both a pedagogic and psycholinguistic level. From a pedagogic perspective, the goal is to provide empirical support to enhanced incidental learning, i.e. learning conditions capable of resulting in unconscious detection and implicit knowledge. As a psycholinguistic corollary of such aim, the study investigates the relationship between the level of consciousness at the point of learning and the kind of knowledge gained.

Among the existing empirical studies, the one carried out by Webb and colleagues in 2013 was chosen as the main refer point for the design, since it combines reading while listening and incidental learning of formulaic sequences. As previously reported in length (§ 3.2.1), Webb and colleagues had their subjects read and listen to a graded reader in incidental learning conditions. The text included a variable number of repetitions of the semantically-opaque target collocations, which were composed of known words. As their study focused on the role of frequency, the number of occurrences of the target items was manipulated through the different experimental groups, which respectively encountered the collocations 1, 5, 10 or 15 times. Participants were tested before the treatment for receptive knowledge of written form and immediately after the treatment for receptive and productive knowledge of form and meaning.

The current study partly replicates Webb and colleagues' design, adapting it to the present research questions and aiming to overcome some of its acknowledged limitations.

As in Webb et al (2013), the participants read and listened to a graded reader. Contrary to the 2013 study, the independent variable here manipulated is not frequency, but instead the kind and degree of noticeability of enhancement devices. Therefore, the number of occurrences for each target idiom is kept constant throughout the experimental groups. Differently from Webb et al (2013), no pretest was delivered. The consideration was made that a pretest would have hinted to the actual scope of the treatment, thus hampering the incidental nature of the desired learning. Moreover, the literature agrees about the importance of delayed posttests for the assessment of knowledge, for which a 3-week delayed posttest was added to the design.

To best of my knowledge, no experimental research has ever been conducted about the acquisition of Italian L2 formulaic language. Therefore, the first research question mainly aims to confirm for Italian idioms the positive findings the EFL literature displays about the possibility to incidentally learn L2 formulaic sequences from reading and reading while listening (Pellicer-Sánchez 2017, Webb et al 2013, Webb & Kagimoto 2009). In order to address this question on an explicit-knowledge level, after the treatment participants carried out offline tests of receptive and productive knowledge of form and meaning, which were designed according to the existing empirical literature (Choi 2017, Gyllstad 2009, Laufer & Girsai 2008, Nguyen & Webb 2016, Peters 2012, Szudarski 2012, Webb & Kagimoto 2009, 2011, Webb et al 2013). The experimental groups exposed to the treatment were compared with a control group performing the tests only.

Implicit knowledge is increasingly recognized as the primary aim for language teaching, with a strong priority over explicit knowledge. It has been theoretically affirmed that focusing on the input meaning is the most likely means for learners to trigger implicit learning and achieve implicit knowledge. However, this statement lacks empirical confirmation, with researchers mostly assuming the implicit nature of knowledge gained in incidental learning conditions such as reading and reading while listening (Rebuschat 2013). Such an assumption needs to be verified through empirical data, this being an acknowledged gap in the empirical literature. This gap is addressed through a self-paced reading posttest (both immediate and delayed), which is recognized to be capable of assessing interiorized knowledge (Keating & Jegerski 2015, Marsden et al 2018, Suzuki 2017).

Given the desirable nature of implicit knowledge, it needs to be pointed out that its acquisition is attested to be extremely slow, and that this results problematic from a language-course perspective. Therefore, it would be highly beneficial for language instruction to find a technique capable of speeding up learning while keeping it implicit. This aim is pursued here by creating the conditions for enhanced incidental learning, i.e. by adding unobtrusive enhancement devices to the first occurrences of the target items in incidental learning conditions. According to the Law of Contiguity (Ellis 2001), conscious noticing of new items needs to take place first, and this allows subsequent statistical learning to occur implicitly. However, in the present work the hypothesis is formulated that unconscious detection can be capable of triggering learning as well. The second research question aims to address this issue, by exposing participants to enhancement devices with growing levels of noticeability. With this goal, the sample was divided into four experimental groups. All of them were exposed to increased frequency of the target items, while for only three of the groups different kinds of enhancement were added to the first two occurrences of the target sequences. In order of presumed noticeability, the chosen enhancement formats were aural enhancement, typographical enhancement, typographical + aural enhancement. The scores in both offline and online tests were compared among experimental groups and with the no-treatment control group.

Finally, the present study aims at contributing to the study of the relationship between levels of consciousness at the point of learning and kind of knowledge gained. Namely, eye-tracking studies demonstrate that more attention to linguistic items (longer and more numerous fixations) results into more learning (e.g. Pellicer-Sánchez 2016). However, the possibility for learners to gain implicit knowledge through conscious attention is strongly debated. Besides, the effects of different kinds of input enhancements on levels of attention have not been investigated yet. The third research question focuses these issues. The learners' level of consciousness is measured at the process level through eye-tracking. In other words, a subsample of the participants, equally distributed in the four experimental groups, carried out the reading-while-listening task while having their eye movements recorded. Furthermore, the whole sample carried out retrospective verbal report, where participants were asked to recall whether they had noticed any enhancement and if they paid conscious attention to learn the enhanced items.

In order to deal with the fourth research question, data about fixations on the target words from the eye-tracking are meant to be triangulated with the scores of the offline and online posttests and with the outcome of the stimulated recalls. This allows to contribute clarifying two main issues: (i) whether incidental learning conditions actually result in implicit learning; (ii) whether an augmented level of consciousness results in explicit knowledge, implicit knowledge, or both.

Summarizing, the research questions formulated for the present study are the following:

1. Can formulaic sequences be learned incidentally through exposure to bimodal presentation of reading passages, without any explicit instruction?

a. If knowledge of the target formulaic sequences is gained, is this knowledge implicit?

b. If knowledge of the target formulaic sequences is gained, is this knowledge explicit?

2. Does adding enhancement to the first two occurrences of the target formulaic sequences affect learning and, if so, is one enhancement format among typographical, aural or both, more effective?

3. What is the level of consciousness at the point of learning in enhanced incidental learning conditions?

4. What is the relationship between level of consciousness at the point of learning and the kind of knowledge gained?

4.2. Participants

Participants were 83 Chinese native speakers with an average age of 20.1 and including 57 females. All of them were enrolled in the Marco Polo – Turandot program, which provides Chinese students with 10-month intensive Italian language teaching (24 hours per week) in order to prepare them to attend Italian universities, conservatories and art academies. Subjects were recruited through their Italian teachers, and they were offered 10 extra points in one of the course assessment as a compensation for their time and participation.

In order to gather a sufficiently large sample sample, three data collections during two subsequent academic years were necessary. The first and the second data collection took place during the 2017/2018 academic year, including respectively 32 students from Università degli Studi di Pavia and 20 students from Università per Stranieri di Siena. The third data collection was carried out during the 2018/2019 academic year and included 31 students from Università degli Studi di Pavia. Only the third-data-collection participants had their eye movements recorded through eye-tracking techniques. The pilot study participants attended the Marco Polo – Turandot program at the Accademia Lingua Italiana (Assisi, PG).

In order to meet ethical requirements for empirical research, the subjects were carefully informed of the experimental procedures through a detailed information sheet and a consent form, both translated in their first language. Namely, the forms provided them with a study summary, information about the voluntary nature of participation, kinds of measurement performed as part of the study, procedures, duration, possible benefits and risks. All participants were informed that anonymity would be preserved in the data collection and storage. Given the incidental and implicit nature of the learning this study intends to observe, the information sheet and the consent form the participants signed before the treatment did not report the actual purpose of the experiment. Indeed, informing the subjects that their learning of FL was monitored would have totally inhibited this learning to be incidental and implicit, thus hampering the core aim of the study. Therefore, the study real objectives were reported in a second consent form the participants were asked to read and sign after taking the delayed posttest. Only the students signing both the first and the second consent form were included in the experiment.

The treatment materials were designed in order to be appropriate for incidental learning of subjects with at least an A2 level of proficiency. Therefore, a CILS (official certification of Italian as a foreign language provided by the Università per Stranieri di Siena) A2 proficiency test was delivered and only the subjects passing it were included in the study. Moreover, subjects with a B2 proficiency might have known the target items prior to the treatment. Therefore, no student with such a level was included in the study. Since delivering two proficiency tests was not viable for timing and logistic reasons, a confirmation about the students not reaching the B2 level was provided by their language teachers. Furthermore, during the retrospective verbal reports all of the subjects were asked whether they had ever seen the target items before, and the data about already-known idioms were discarded.

Finally, since incidental learning of collocational properties and meaning from the context assumes a high understanding of the input meaning, data from students scoring less than 75% in a comprehension test about the treatment text were discarded.

To sum up, the selection criteria implemented in order to create the final pools for the three data collections were the following:

a. only Chinese native speakers attending the Marco Polo - Turandot program were included;

b. only subjects with an attested proficiency level between A2 and B1 were included;

c. only subjects who during the retrospective verbal reports confirmed not to have seen the target items ever before were included;

d. only subjects who scored 75% or more in a comprehension test about the input meaning were included;

e. only subjects who had signed both the first and the second consent form were included;

f. only subjects who participated in all the phases of the experiment were included.

Keeping into account the CILS proficiency test scores, participants were randomly assigned to 4 experimental groups and one control group. Randomization took place blocking on L2 proficiency, i.e. proficiency differences among the groups did not to reach significance. The four experimental groups were exposed to four different treatments (see infra, § 3.4), while the control group only performed the posttests.

The five different groups were created during each of the three data collections. All of the groups were part of comparable populations as of age, L1, linguistic background, L2 proficiency. Furthermore, they followed exactly the same procedures and were exposed to the very same treatment and test materials. Therefore, merging the data from the three data collection considering the three samples as a single pool is not considered problematic.

The final composition of the sample is illustrated in table 4.1.

Table 4.0.1. Number of participants per group in each data collection

	1st data collection	2nd data collection	3rd data collection	Total
Group 1	7	4	7	18
Group 2	5	5	7	17
Group 3	6	5	6	17
Group 4	9	2	4	15
Group 5	5	4	7	16

4.3. Target structure: formulaic language

Formulaic sequences were chosen as a target structure for the present study for different reasons. First, they are widely recognized as both an essential and a problematic part of L2 acquisition. As section 2.1 reports in length, importance and difficulty of FL can be observed and confirmed from various perspectives. From a statistical point of view, formulaic sequences are both extremely frequent and highly dispersed in the language. Their structural features make them hard to recognize and master for L2 speakers, while on a pragmatic level, correctly interpreting the appropriateness of a formulaic phrase is a task even highly proficient L2 speakers tend to fail or avoid. Moreover, on a psycholinguistic level, employing formulaic language implies a processing advantage for native speakers which would have a strong impact on L2 learners' online language use, and which is therefore highly desirable.

On a different level, collocational properties of words are considered as the aspect of vocabulary knowledge which responds the most to statistical learning, and which is more likely to be stored in the implicit memory (Nation 2001). Given the present study's focus on incidental learning and implicit knowledge, this feature makes formulaic language the correct target structure.

4.3.1. Idioms: theoretical and methodological rationale

Among the numerous kinds of formulaic sequences taxonomies can identify, the choice was to focus on idioms. The rationale of such a choice is twofold, i.e. it presents both theoretical justifications and methodological reasons. As compared to other kinds of formulaic sequences, idioms present additional factors of difficulty, mainly related to their non-compositionality (§ 2.1). On a different level, idiomatic expressions are worth focusing on because despite the existence of scientific literature about L1 processing of Italian idioms (e.g Tabossi et al 2009), to the best of my knowledge no study of L2 acquisition of Italian idioms has ever been performed.

Some of the linguistic features making idioms a valuable field of investigation also provide strong methodological advantages, capable of overcoming part of the limitations claimed in existing empirical research.

The first consideration on this regard is related to the choice of having target FSs made up of either known or unknown words. The decision was made to have experimental subjects who knew the meaning and the form of the components, but were not familiar with the whole multi-word unit. This condition is the most common in the literature (e.g. Choi 2017, Durrant & Schmitt 2010, Gyllstad 2009, Laufer & Girsai 2008, Nguyen & Webb 2016, Peters 2012, Webb & Kagimoto 2009 and 2011, Webb et al 2013), and can be justified with a number of claims. First, known words eliminate factors such as pronounceability, orthography, morphology which may affect single-item word difficulty (Laufer 1997).

Second, participants working with unfamiliar FSs containing unknown words would face a double learning task, as they would need to learn both the FS and its components. In other words, their attention would be diverted from the compositional properties to deriving the meaning of single words (Webb et al 2013). Therefore, the posttest measurements would be affected by factors such as single-word learning, which is not the focus of the present study. At the same time, the impact of the pedagogic techniques aimed at FS learning might be lessened, as Szdudarski and Carter (2016) claim in the limitation section of their paper, when explaining possible reasons for a lack of frequency effect on the acquisition of infrequent collocations (§ 3.2.1). Third, the most common condition an L2 speaker faces is the need to learn FSs made up of high-frequency, known words. Therefore, recreating the same situation in the experimental conditions improves the ecological validity of the study.

Given the need for known components of the FSs, the issue of measurements rises. According to Webb and colleagues (2013), one of the main reasons the acquisition of FSs is under-researched relates to this point: most vocabulary studies measure knowledge gains as of form-meaning connections, but assessing this kind of learning is not straightforward in the case of FSs. Indeed, the knowledge of the single components meaning automatically implies knowledge of the FS meaning, in case of semantically transparent FS. This in turn implies the impossibility to assess the effectiveness of pedagogic treatments as of the learning of FS meaning. Keeping these considerations into account, the choice was made here to have idioms as target items, since figurative meaning ensures semantic opacity and therefore allows measuring improvements in the knowledge of FS meaning.

A factor in need to be taken into account when defining the features of the target idioms was the necessity of avoiding a pretest, as recommended in Webb and colleagues' (2013) work. A first reason not to have participants take a pretest is to avoid learning effects. A second and more crucial reason is related to the very goal of the present study, i.e. measuring incidental and implicit learning. In order for this to be possible, participants needed to be unaware of the actual target of the treatment performed, and they were told the only tests were going to be about the general understanding of the reading text. A pretest would have provided the participants with clues about the aim of the treatment, therefore hampering such incidental conditions and increasing the likelihood of intentional, explicit learning. The requirement to avoid a pretest is reported and met in experimental literature about incidental learning of vocabulary, and the methodological solution often adopted is to employ pseudo-words (e.g. Pellicer-Sánchez 2016; 2017). However, the choice was made here to give priority to ecological validity, which brought to the selection of low-frequency idioms.

Summing up, the target items in the present study were selected in order to be:

a. composed of high-frequency, known words;

b. semantically opaque (idiomatic expressions with high non-compositionality)

c. arguably unknown to the participants before the treatment (low frequency)

4.3.2. Selection of the target idioms

In order to select idiomatic expressions meeting the 3 requirements reported above, a number of steps were necessary.

First, in order for the findings of the present study to be more generalizable, the target idioms need to meet the statistical definition of collocation. Therefore, a first selection of idiomatic expression was drew from the Dictionary of Italian Collocations (http://www.dici-a.it), an online tool gathering Italian collocations selected through rigorous statistical procedures (Spina 2010). The website also allows to filter the levels of proficiency learners knowing the collocation are supposed to belong to. The first selection of potential target items included Italian collocations with idiomatic meaning (condition b) from the B2 proficiency band (condition c). 32 idioms met these requirements.

The 32 FSs went through a further selection in order to meet condition (a) above, i.e. only idioms composed of high-frequency, known words were kept in the list. With this aim, three verifications were carried out on the single-word components. First, only words occurring among the first 2000 in Italian frequency lists were kept. Second, only words reported as part of an A2 vocabulary were kept (Spinelli & Parizzi 2010). Third, a list of the words was presented to the subjects' Italian teachers in order to make sure they were all part of the participants' syllabus, and only the words marked as 'known' by the teachers were kept. 13 idioms met these requirements.

According to condition (c), the idioms had to be infrequent enough to be considered unknown to the participants without the need of a pretest. Therefore, they were chosen in the B2 proficiency band in the dictionary of Italian collocations. In addition, a 1 million 8 thousand learner corpus was queried and only the idioms with no occurrence in the corpus were kept in the list. 10 idioms met this requirement. Moreover, during the pilot study and the retrospective reports of the main data collections participants were asked whether they had ever seen the idioms before, and the data relative to idioms with a positive response were discarded.

Finally, an L1 Italian Chinese interpreter and an L1 Chinese speaker checked together the final list in order to exclude the existence of corresponding figurative meanings in the participants' L1.

The final list of target idiomatic expressions is reported in table 4.2. It includes 10 idioms, of which 5 are composed by noun + adjective, and 5 by verb + object.

Table 4.0.2. Target idioms

Idiom	Composition	Literal meaning	Figurative meaning

(C'è) una bruttta aria	Adj + Noun	There is a bad air	There is a bad atmosphere / things are getting nasty
Aria fritta	Noun + Adj	Fried air	Nothing important or concrete
Doccia gelata	Noun + Adj	Cold shower	Unpleasant suprise
Testa calda	Noun + Adj	Hot head	Impulsive person
Braccio destro	Noun + Adj	Right arm	Right hand man
Aprire gli occhi	Verb + Obj	To open one's eyes	To realize something
Costare un occhio	Verb + Obj	To cost an eye	To be very expensive
Perdere la faccia	Verb + Obj	To loose one's face	To loose face
Toccare il cielo	Verb + Obj	To touch the sky	To be very happy
Mettere il naso	Verb + Obj	To put one's nose in	To stick one's nose

4.4. Instructional material

The aim of the present research is to verify the effectiveness of enhanced incidental learning conditions. As already discussed in length (§ 3.5), enhanced incidental learning consists in two phases: in the first phase, a first memory trace of the unknown item is created. Therefore, pedagogic intervention should increase salience for the first occurrences of the target items. In the second phase, implicit, statistical tallying takes place, which requires from the instructor to set incidental learning conditions.

The treatment designed responds to these considerations. Subjects were exposed to incidental learning conditions, i.e. reading while listening to a graded reader. In order to boost and direct incidental learning, two strategies were adopted:

(i) the first two occurrences of the target items were enhanced, with the aim of favoring detection of the new forms, which in turn could make subsequent statistical learning possible. In order to investigate the role of awareness in this first learning phase, the format and noticeability of the enhancement devices employed were manipulated as an independent variable.

(ii) After the first two occurrences, it is assumed that a memory trace of the target items should have been created and the possibility for statistical learning triggered. In order to favor this process, all enhancement devices but increased frequency are removed.

The present section describes in detail the design of the instructional material.

First, when creating the reading material for the study, a crucial point was to define the number of repetitions for the target idioms. The literature review (§ 3.2.1) showed that frequency of occurrence has a clear role in vocabulary learning. However, there is less agreement about the minimum number of repetitions necessary to trigger learning. Combining the existing empirical findings with feasibility factors, 7 occurrences of each of the ten target idioms were embedded in the reading-while-listening text. As mentioned above, vocabulary studies report learning effects starting from 2 repetitions, while Webb and colleagues (2013) found FL knowledge improvements with 10 repetitions or more, in a text including 18 unknown items and 5 thousand words. The present study tests 10 unknown items in less than 5 thousand words (see infra), therefore 7 repetitions per item are considered enough to define an input-flood condition where learning is highly probable.

The total length of the text was defined as a function of the lexical coverage that is necessary for learners to infer meaning from context. The exact percentage is debated, but researchers tend to estimate that between 95% and 99% of the words need to be known in a text for subjects to understand and acquire the meaning of new items (Nation 2001, 2006). As the single words were known while the figurative meaning of the whole FS was not, in the present study each of the ten target idioms was considered as one new-meaning item in need to be inferred. Each target idiom occurred 7 times, so the unknown items in the text were a total of 70. A total length of at least 3500 words was thus calculated to be necessary, in order for the unknown items to constitute no more than 2% of the text. Plausibility and distribution considerations brought the text to a total of 4700 words, the target unknown items thus constituting 1,5% of the text.

A 7-chapter thriller story ("La ricetta segreta"[1]) published for an A2-proficiency target was modified and adapted in order for each chapter to contain all of the ten target idioms in informative and plausible contexts. This allowed a controlled and homogeneous distribution of the FSs.

As mentioned, all of the words composing the story needed to be known to the participants, except for the target idioms. The text was therefore scanned by a software (www.corrige.it), which signals through graphic enhancement (bold, italics, etc.) the frequency band each word belongs to. Only words among the 2000 most frequent were kept. Less frequent words were either replaced or presented together with

[1] The secret recipe

L1 glosses or images as a means of assuring understanding. A total of 16 L1 glosses and 7 images were provided, obviously with none of them referring to the target items.

When it comes to additional enhancement for the first two occurrences, it has been pointed out (§3.1; 3.2) that the choice was made to employ and combine typographical enhancement and aural enhancement.

Despite the mixed results, empirical evidence reported in the existing literature supports input enhancement effectiveness as of FL learning (§ 3.2.1). However, some gaps are evident and need more empirical data. First, even though implicit knowledge is widely accepted as the main aim for language teaching, only two studies (Sonbul & Schmitt 2013; Toomer & Elgort 2019) carried out implicit knowledge assessments. Moreover, while different researchers verified effects of input enhancement on the participants' level of consciousness through eye-tracking when teaching grammar structures, no study did the same when FL learning was concerned. Recording such measures and comparing them with the kind of knowledge resulting in the posttests (implicit or explicit) is crucial for a deeper understanding of the relation between levels of consciousness and learning. This understanding can in turn be a foundation for effective, research-informed pedagogic techniques capable of resulting in the acquisition of implicit knowledge. The hypothesis is here made that the initial image of new item can be stored not only as a result of conscious noticing as usually maintained, but also through unconscious detection. In order to test this hypothesis, different formats of input enhancement with growing levels of noticeability are employed.

With the aim of contributing to clarify the discussed gaps, four versions of the reading-while-listening text were created, including a combination of three different kinds of input enhancement.

- Version 1: increased frequency + typographical enhancement (bold).
- Version 2: increased frequency + aural enhancement (0.4 seconds pause before and after the target items. Pauses were digitally added to the same audio file used for the version 1 text, in order to avoid any other variation in the aural input).
- Version 3: increased frequency + typographical and aural enhancement (bold + 0.4-seconds pause before and after the target items).
- Version 4: increased frequency only.

Following a between-group design, each version of the text was read and listened to by a different experimental group (table 4.3).

Table 4.0.3. Treatments and groups

Group	n	Treatment
1	18	Increased frequency + typographical enhancement (TE)
2	18	Increased frequency + aural enhancement (AE)
3	17	Increased frequency + typographical and aural enhancement (TAE)
4	15	Increased frequency only (IFO)
5	16	No treatment – control group

4.5. Dependent measures

The different tests employed reflect the need to address the gaps in the existing literature, as defined in the research questions.

First, individual differences among participants needed to be taken into consideration for the data analysis and the creation of the experimental groups. Therefore, a proficiency test (CILS) and two working memory tests (digit span and operation span) were carried out before the treatment.

One of the main goals of the present study is to investigate the relation between input enhancement, levels of consciousness at the point of learning and kind of knowledge gained. In order to measure the level of consciousness at the process level, participants' eye movements were recorded while reading and listening to the treatment text. Moreover, retrospective verbal reports were collected (RQ 4).

Nation (1990; 2001) identified several aspects of vocabulary knowledge, including among others form, meaning and collocations. Different scholars (N. Ellis 1994, R. Ellis 2004, Schmitt 2008, Sonbul & Schmitt 2013) affirm that some of these aspects are more likely to be learned intentionally and to remain part of the explicit knowledge, while other aspects of knowledge are probably largely implicit. Namely, components related to semantics, form, meaning and their connections are best learned explicitly, while aspects such as collocational properties and frequency intuitions can be best learned implicitly. These considerations are here taken into account when designing tests for explicit and implicit knowledge of the target idioms. The three offline tests aiming to measure explicit knowledge focus on form and meaning. On the other hand, the online test meant to assess implicit knowledge (self-paced reading) addresses collocational knowledge.

As mentioned above, no pretest took place, in order not to provide participants with clues about the actual aim of the study, since this could undermine the incidental nature of learning. Participants carried out the immediate posttests immediately after the treatment, while the delayed posttests took place 3 weeks later. The delayed posttest was identical to the immediate posttest except for the order in which the test items were presented.

4.5.1. Proficiency and working memory tests – individual differences

There is evidence that higher proficiency (Lee & Pulido 2017) and larger vocabulary knowledge (Murphy, Miller & Hamrick 2018, Vilkaite 2017) positively affect learning of vocabulary and collocational properties. In order to account for individual differences as covariates in the statistical analysis, the participants' level of proficiency was tested before the experimental treatment. No validated vocabulary test exists for Italian, therefore proficiency scores are used here as a function of vocabulary knowledge too.

The instructional material and the target idioms are designed and selected to be suitable for learners with proficiency levels from A2 to B1 (§ 5.2, 5.3). Therefore, an A2-level Italian certification test was carried out. Four official certifications exist for Italian as second language, provided by four different institutions, and namely: PLIDA (Società Dante Alighieri), CELI (Università per Stranieri di Perugia), RomaTre (Università di RomaTre) and CILS (Università per Stranieri di Siena). CILS was chosen here, being the most familiar for Chinese learners. Subjects not reaching the 50%+1 of correct answers, i.e. not passing the level test according to the CILS policy, were excluded from the study. They carried out the treatment with their classmates, but their data were discarded from the analysis. Scores from learners passing the test were included in the statistical analysis.

Students with a level higher than B1 would have risked knowing the target idioms before the treatment. It was chosen not to rule out this possibility by carrying out a B2 proficiency test. Indeed, B2 tests are highly time-consuming (3 hours) and the likelihood of finding a B2-level learner was less than minimal according to the class teachers. Moreover, during the retrospective verbal reports (see infra), the subjects were asked whether they had ever met the target idioms before the treatment, which allowed discarding the (rare) data relative to already-known formulaic sequences.

Working memory (WM) refers to a set of cognitive processes involved in the processing, storage and retrieval of information (e.g. Beddeley & Hitch 1974). Beddeley's model of WM includes a short-term storage component and an attentional control component called central executive. WM is a capacity-limited system, i.e. the amount of information that can be actively maintained in the focus of attention is finite. Since it is the central executive that manipulates the contents of WM, this part has been recognized as the main determiner of individual differences in WM (e.g. Engle 2002). Experimental evidence converges about the significant role of working memory for language learning and acquisition, which is why it is here included among the individual differences measures to be taken into account

(Engle et al 1992, Foster et al 2015, Jullfs & Harrington 2011, Kormos & Sàfàr 2008, Linck & Weiss 2015, Malone 2018). For logistical reasons, it was only possible to collect WM data during the third and last data collection.

In the present study two measures of WM are employed. The first is a digit span task, i.e. a simple WM span task. The second is a modified version of the operation span task, which represents a more complex measure of WM processing (Foster et al 2015). Both tests were computer-supported, and they were created on the same Paradigm software employed for the self-paced reading test.

Digit span is among the oldest and most widely used neuropsychological tests of short-term verbal memory (Richardson 2007). In the digit span task, subjects read lists of digits and are then requested to rewrite them. The list length progressively increases, so that the first sequences participants are requested to memorize and rewrite are comprised of 3 digits, the following sequences include 4 digits and so on until a maximum length of 9 digits. Three trials are presented for each list length (Woods et al 2011). If a subject fails all of the trials of a certain length, the test ends. The total number of lists correctly reported constitutes the digit span score for each subject.

The operation span (Ospan) is a more complex task, because a distractor task is added to the memory task. The items the subjects need to remember are letters, and simple math problems are used as distractors. Subjects first see a letter, then they need to state whether a simple math equation is correct or not. Then again, they see a letter and an equation and so on. Such letter-equation sequences are repeated from 3 to 7 times for each trial, with an unpredictable length each time. Each length trial occurs two times in random order. After each trial, the subjects are required to recall the letters they saw, in order. The final score is calculated summing the number of letters correctly recalled. If more than 20% of the math equations are not answered correctly, however, the trial is discarded regardless of the correct letters.

4.5.2. Comprehension test

Inference of unknown meanings from context is only possible if 95%-99% of the words are known and the general comprehension of the text is high (Nation 2001; Schmitt 2008). The reading-while-listening text is designed in order to fulfill such condition, but a further verification was considered necessary. Therefore, a general comprehension test about the meaning of the text was delivered right after the treatment. The test was comprised of 2 exercises. In the first one, subjects were required to put in the right order 6 images picturing salient points in the story. The second exercise asked 14 T/F questions. Students scoring less than 75% in this test were excluded from the study, i.e. their data were discarded from the analysis.

The general comprehension test also played a crucial role from the incidental-condition learning point of view. Indeed, participants were told at the beginning of the treatment that the activity main goal was to exercise their comprehension of Italian written texts. They were told to focus on the meaning because that was going to be assessed in the following test. Therefore, delivering the announced general understanding test was considered appropriate.

4.5.3. Eye-tracking – level of consciousness at the process level

According to the eye-mind hypothesis (Just & Carpenter 1980), tracking people's eye movements allows determining the direction and amount of their attention. This statement relies on two main assumptions (Pickering et al 2004). First, what is being fixated is what is being considered, i.e. readers try to interpret words as they are encountered. Second, the time spent fixating an item reflects the amount of cognitive effort necessary to process it. Therefore, eye-tracking (ET) is capable of providing the researcher with information about the subjects' level of attention at the process level. Such information is especially reliable because it results from a direct measure of the processing effort, i.e. eye movements are automatic and elude conscious control. At the same time, ET measures do not imply the output of any additional task such as decisions, recall of production, which may be affected by strategy or other factors. In addition, eye-tracking presents further advantages, such as a limited variance due to individual differences, temporal precision and the possibility to engage with reading in a natural way (Conklin, Pellicer-Sánchez & Carroll 2018).

Given its features, eye-tracking has recently spread in psycholinguistic research about reading and it is currently considered as a desirable methodological standard.

A number of studies employed eye-tracking in order to investigate L2 vocabulary and FL processing (Carroll et al 2016, Siyanova-Chanturia et al 2011, Yi et al 2017) and learning (Choi 2016, Godfroid et al 2013, Mohamed 2017, Pellicer-Sánchez 2016). However, no study compared the information about consciousness at the point of learning provided by eye-tracking measures with scores of both offline and online posttests. This triangulation is performed in the present study, with the aim of investigating the relation between attention at the point of learning and kind of knowledge gained (for a more detailed review of the literature, see chapters 2 and 3).

When reading, the eye stops to process information and then makes a rapid movement to the next location where information is available. The intervals when the eye stops are called fixations, and it is during fixations that the cognitive system perceives and processes information. The movements from a fixation to the next one are called saccades, and during a saccade the eye moves so quickly that no visual information can be obtained. However, the processing of already perceived information can continue. 10 to 15% of the times, saccades bring the eye backwards to already-encountered sections of the text. These movements are called regressions. Eye-tracking technology allows to measure, for pre-defined words or word sequences (region of interest, ROI), the number and duration of fixations and regressions and the number and length of saccades. Broadly, longer and more numerous fixations as well as more regressions on a region point to a harder cognitive effort. On the contrary, items and regions with shorter fixations or skipped at all are easier to process. More in detail, ET measures are classed into three typologies, which are believed to provide information about different stages of processing:

- Early measures reflect highly automatic word recognition and lexical access processes. They include:

o Skipping rate: the proportion of words that are not fixated at all during first pass reading. Skipped words are supposed to be processed during the fixation of the previous word, and factors such as frequency, lexical status (function or content words) and predictability can determine whether a word is skipped.

o First fixation duration: the length of the first fixation on a word or on a region of interest (ROI).

o Gaze duration (first pass reading time): all fixations made on a word or ROI before the gaze exits. Both first fixation duration and gaze duration are indexes of lexical access, and therefore they are affected by frequency, familiarity, meaning ambiguity, predictability and semantic association.

- Late measures tend to reflect more conscious and controlled processes, and they are influenced by higher-level variables such as context, sentence or discourse. They include:

o Total reading time: the sum total of all fixations made on a word or ROI, including first fixation and re-readings.

o Fixation count: total number of fixations.

- Intermediate measures include regression measures, which are hard to classify as either early or late, since they can point to both the difficulties met when first meeting an item and to the later attempts to overcome such difficulty. They include:

o Regression path duration: time spent on the ROI itself and any prior part of the sentence before moving past the critical region to the right.

o Regression rate: proportion of trials with a regression.

Experimental studies usually focus their analysis on part of the listed measures, according to their aims and target structures. In the present study, the goal is twofold: first, measuring learning effects reflected by fixations and skipping rates changes along the seven occurrences of the target idioms. Second, measuring the effects of input enhancement and enhanced incidental learning conditions on eye movements, by comparing the experimental groups. With these aims and according to the two main studies investigating FL with ET (Carrol et al 2016, Siyanova-Chanturia et al 2011), two ROIs are defined for each idiom: the whole idiom and the final word of the idiom. For the whole idiom, first pass reading time, total reading time and fixation count are studied. Such measures are analyzed in order to investigate lexical access to the idiom meaning and context, as well as the effects of input enhancement. For the final word of the idiom likelihood of skipping, first fixation duration, gaze duration and total reading time are analyzed. Indeed, the more the idiom becomes familiar, the more predictable the second word should be and therefore learning is expected to manifest with increased skipping rates and fewer and shorter fixations.

The reading-while-listening text created for the experiment was adapted to the ET technology copying it on timed slides, which automatically changed synchronized to the audio input, so that the subjects did not have to go through them manually. The text was presented in Courier New font, which is usually preferred because all letters take up the same amount of horizontal space. Font size was 18 and lines were triple-spaced. Care was taken in order for ROIs never to appear as the first or last words of sentences and lines. The reason is to avoid wrap-up effects (longer fixations on the final word of sentences as the whole meaning is recollected) and gaze instabilities that often occur at the beginning of lines. The audio input was provided via wireless headphones, but the participants did not have control over the audio file, which was managed by the researcher.

The eye-tracking part of the experiment took place individually, in a quiet room of a laboratory at Università degli Studi di Pavia. The eye tracker was a GazePoint GP3 HD, which consists in a desk-mounted camera with a sampling rate of 150 Hz and 0.5-1.0 degree of visual angle accuracy. Participants were seated on an adjustable chair at a 40cm distance from a 1680x1050 widescreen monitor. Calibration of the eye tracker was performed per each participant before starting the reading while listening treatment. As reported in the procedure section below, two breaks were allowed during the treatment in order to compensate and limit fatigue effect. After each break calibration was repeated.

4.5.4. Self-paced reading test – implicit knowledge

One the main aims of the present study is to verify gains in the learners' implicit competence through the use of online assessment (RQs 1, 2 and 3). Indeed, numerous studies measured the knowledge created by incidental learning of vocabulary and collocations, but the nature of this knowledge is yet to be investigated (Rebuschat 2013). As a means of testing implicit knowledge, a self-paced reading task was created.

Marsden and colleagues (2018) effectively define self-paced reading (SPR) as "an online computer-assisted research technique in which participants read sentences, broken into words or segments, at a pace they control by pressing a key." (p. 1). A computer software (in this case, Paradigm) measures the time elapsed between each press, i.e. the time the subject spends on each segment. This time is called reaction time (RT), it is usually measured in milliseconds (ms) and it constitutes the main dependent variable provided by the self-paced reading test. The rationale of analyzing RTs lies in the premise that cognitive processes take time, and therefore observing how long it takes subjects to respond to stimuli allows inferences about the mechanisms involved in language processing (Jiang 2012). As clearly stated by Lackman and colleagues (1979), "time is cognition" (p. 133). Namely, a longer RT in one condition than in another is likely to reflect a higher degree of difficulty or a higher level of complexity in operations. According to the anomaly detection experimental paradigm, longer RTs for violations than for correct structures may show implicit sensitivity to errors, i.e. the existence of an implicit competence regarding the structure in exam (Keating & Jegerski 2015).

Different features of self-paced reading enable to be reasonably sure that such task taps into implicit competence, with a minimal involvement of explicit knowledge, if any. First, the self-paced reading task takes place online, i.e. while the comprehension is ongoing, and subjects are required to read as fast as possible. This emphasis on the speed of performance and the transient nature of input display make it unlikely for learners to use linguistic knowledge consciously in such a short time, which usually amounts to a few hundred milliseconds per segment (Suzuki 2017). Second, the receptive nature of the task does not require any production from the subject, and therefore removes a further reason to tap into explicit knowledge. Third, in a well-designed self-paced reading task participant read the sentences focusing on meaning, due to comprehension questions following each critical item. Therefore, they have no reason to pay conscious attention to language form. Additionally, subjects should not be aware of the linguistic structure the test addresses, which is achieved adding filler items. A recent experiment by Suzuki (2017) corroborates SPR capability of assessing implicit knowledge. Suzuki aimed at demonstrating that some of the tests often employed to measure implicit knowledge actually draw on automatized explicit knowledge. Statistical analysis of the scores from 100 Japanese learners with first language Chinese confirmed that tests such as timed grammaticality judgment cannot measure implicit knowledge, despite the time constraint. On the contrary, SPR proved to be an effective assessment tool for implicit competence.

Among the advantages for using SPR, it is noteworthy to cite its relative ease of administration and its ecological validity. SPR leaves control over the reading time to the participant, as in natural reading. The need to press a button to see new segments imposes an additional task, which is often quoted in the literature as a possible interference with the reading naturalness. However, recent empirical data demonstrate a negligible effect of the button-pressing task, which further confirms SPR ecological validity (Suzuki 2017).

In the current study, the SPR test consists of 20 sentences, presented in a moving-window mode. In the moving-window condition, in the first screen all of the words are substituted with dashes. Pushing a button on the keyboard, the participant allows the first phrase to appear. With a second press, the following phrase on the right is made visible, while the first phrase is replaced by dashes, and so on. The moving-window condition is currently preferred to the alternative options of cumulative condition and stationary window. In the cumulative mode the words or phrases on the left remain on the screen when the button is pressed and the new segments appear. As a result, in the end the whole sentence is visible at once. This mode poses two main risks, capable of undermining the data reliability as of online measurements. First, the participant may press the button very quickly in order to have the complete sentence and read it only in the end, not paying actual attention to the single segments as they appear. Second, the participant may re-read the sentence once it is completely visible, thus boosting his/her comprehension above the level of an online-only read. In the stationary-window condition, segments are presented in the center of the screen with the subsequent word replacing the preceding one. All of the words appear in the same position, and no clue is given about the sentence length. Therefore, this reading mode is the most distant from natural reading and has a low ecological validity. Just, Carpenter and Wooley (1982) compared the data collected in the three SPR conditions with eye gaze duration data, and demonstrated the moving-window mode to be the one correlating best with eye movements. For this reason, the moving-window condition is adopted in the present study, and in most of the empirical literature employing SPR (Jiang 2012).

The sentences making up the test are segmented with a phrase-by-phrase presentation instead of the more common word-per-word presentation. The rationale relies on the need to include a whole idiom (or its violation version, see infra) in a single segment, in order for the data interpretation to be clearer. If the target idioms were the only segments to be longer than one word, participants might get a clue about the actual aim of the test, which would have hampered the possibility to collect data about implicit knowledge. This justifies the need of segmenting experimental sentences phrase-per-phrase.

Of the 20 sentences constituting the test, 10 contain a target idiom, and 10 constitute filler sentences containing other lexical items from the reading text. Namely, the fillers include some of the less frequent words occurring in the text, which needed L1 glosses or images in order to the surely understood. In this way, even if participants do understand their vocabulary comprehension is being tested, it is not possible for them to make clear which words or features are under exam.

After each item, subjects are required to answer a true/false question about the general comprehension of the sentence. These questions are crucially meant to keep the participant focus on meaning and not on the form of language, in order to foster the chances for data to reveal information about implicit competence. Therefore, the questions do not address the idioms meaning, and can be answered without focusing or understanding it (Jiang 2012, Keating & Jegerski 2015, Marsden et at 2018). Subjects with a total accuracy score below 70% were excluded from the study, and RTs relative to not-properly-understood sentences were discarded from the analysis.

The aim of the SPR test in the current study is to highlight implicit competence about the target idioms. As reported above, implicit knowledge of vocabulary is likely to relate to collocational properties more than to form-meaning connections. Therefore, the anomaly detection paradigm is here adapted in order to highlight such level of knowledge. Bordag and colleagues' (2015) SPR test is here referred to as a methodological model. Their subjects carried out a self-paced reading task where nouns were combined with either compatible or incompatible adjectives. Researchers measured the difference in RTs between the two conditions, in order to gain information about the subjects' mental lexicon. In the present experiment a similar design is employed. The independent variable is manipulated into two levels, i.e. idioms are presented to the subjects either in a correct form or with a violation. The correct form replicates the one participants were exposed to during the reading-while-listening treatment. As of the violation, one of the two words forming the idiom is replaced by a different one, which is chosen in order to be semantically plausible, equally known to the subjects and comparable to the original word in terms of length and frequency. Also, the distracters are different from those employed in the form-recognition offline test.

Example (4.1):

Mette il naso – mette i piedi*

*He puts his nose - he puts his feet**

The design and counterbalancing of the experimental sets was the result of a multiple-step validation process. The first version of the test to be piloted included 20 experimental items, i.e. each subject was exposed to both the correct form and the violation of each idiom. The sentences providing context for the two conditions were different though comparable in terms of length and comprehensibility. Two sets of sentences were created, so that the context sentences associated with the correct idiom in the first set contained the violation in the second set and vice versa. This design had the advantage of providing 10 target RTs per subject.

Example (4.2):

SET 1

Carlo vuole sapere tutto di me e mette il naso fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his nose in my stuff. I don't like him!

Se vuoi informazioni su Claudio, devi chiedere a Chiara. Lei conosce molte persone, e mette i piedi* dappertutto.

If you want information about Claudio, you should ask Chiara. She knows many people and puts her feet everywhere.*

SET 2

Carlo vuole sapere tutto di me e mette i piedi* fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his feet in my stuff. I don't like him!*

Se vuoi informazioni su Claudio, devi chiedere a Chiara. Lei conosce molte persone, e mette il naso dappertutto.

If you want information about Claudio, you should ask Chiara. She knows many people and puts her nose everywhere.*

However, a pilot data collection (see infra, § 5.7) showed the test not to be effective in measuring differences between RTs of correct idioms and violation. Given the non-normal distribution of data, a nonparametric test was carried out comparing RTs at the correct idiom and at the violation. The related-samples Wilcoxon Signed Rank test showed no significant difference (p=.7). As reported in the literature (Jiang 2012) such result can be affected by the differences in the sentences providing context. Therefore, a second pilot was run, testing a pool of 12 native speakers. In this case, the same sentence provided context for both the correct form and the violation, and each subject was exposed to both. Two sets were created so that the position of the correct and the violation condition switched between the sets. In other words, if the correct condition occurred first in set 1, it occurred second in set 2 and vice versa. Moreover, care was taken in having at least 10 items between the two occurrences of the same sentence.

Example (4.3):

SET 1

Item (5) Carlo vuole sapere tutto di me e mette il naso fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his nose in my stuff. I don't like him!

Item (17) Carlo vuole sapere tutto di me e mette i piedi* fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his feet in my stuff. I don't like him!*

SET2

Item (5) Carlo vuole sapere tutto di me e mette i piedi* fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his feet in my stuff. I don't like him!*

Item (17) Carlo vuole sapere tutto di me e mette il naso fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his nose in my stuff. I don't like him!

The same statistical analysis carried out for the pilot study data highlighted no significant difference between conditions, i.e. native speakers read correct idioms and violations at the same speed ($p=.6$). Rather, RTs were significantly shorter for the second occurrence of the idiom, notwithstanding the experimental condition, as highlighted by a Wilcoxon Signed Rank test ($p=.003$) carried out comparing RTs in the target segment for the first and the second occurrence.

Such results are most likely caused by a repetition effect, which implies an unnatural way of reading (superficial in this case) due to having already seen the sentence before. In order to avoid this effect, the choice was finally made to create two presentation lists, one for each level of the independent variable. This way, each subject was exposed to each item in only one experimental condition, i.e. participants saw each idiom either in its correct form or containing a violation, according to the counterbalancing recommendations reported in the literature (e.g. Keating & Jegerski 2015). Therefore, each set includes 10 target sentences, of which 5 contain the correct idiom and 5 contain a violation. Idioms appearing in the correct condition in set 1 present the violation condition in set 2 and vice versa, but apart from that the sentences are identical. Before starting the actual test, subjects go through 5 practice trials.

Example (4.4):

SET 1

Carlo vuole sapere tutto di me e mette il naso fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his nose in my stuff. I don't like him!

SET 2

Carlo vuole sapere tutto di me e mette i piedi* fra le mie cose. Non mi piace!

Carlo wants to know everything about me and he puts his feet in my stuff. I don't like him!*

Comprehension question: I think Carlo is nice.

The two sets are evenly distributed in the experimental groups, so that a similar or identical number of subjects from each group carry out each set. This kind of design halves the quantity of data collected, because subjects are tested on 10 target items instead of 20. However, it is still preferable according to the methodological literature, and the validation test carried out with a new sample of 16 native speakers confirmed it. The related-samples Wilcoxon Signed Rank test comparing RTs to violations and correct idioms showed a significant difference between them ($p=.001$), which validates the test and allows to expect significant results from the main experiment NNS participants.

4.5.5. Offline tests – explicit knowledge

Offline tests are aimed to measure knowledge of both meaning and form at both the productive and receptive level (Nation & Chung 2009, Webb et al 2013, Webb & Kagimoto 2009). In other words, four levels of vocabulary knowledge are assessed, and namely: form recall, meaning recall, form recognition and meaning recognition.

For the present study, this current practice has been applied, albeit adjustments related to the target items' features were necessary.

The following three posttests assessed the explicit knowledge gained by means of the experimental treatment (RQs 1 and 3). The tests were carried out in pencil and paper format, and when working on each test participants did not have access to the other two. Since the aim was to measure explicit knowledge, no time constrain was imposed (Ellis et al 2009).

The productive test was administered before receptive tests, in order to minimize test-related learning effects. Stubbe (2019) empirically demonstrated that if participants take receptive tests before productive tests, outcomes of the latter are highly affected. On the contrary, taking productive tests before receptive tests does not have significant effects on the results. These findings confirm previous studies by Laufer and Goldstein (2004) and Laufer and McLean (2016), which argued productive assessment to be more difficult than receptive assessment. Keeping all this into account, the tests were administered in the following order:

1) L1 to L2 translation: form and meaning recall, i.e. productive knowledge of form and meaning.

The test comprised 10 items, each consisting of a Chinese sentence and an Italian equivalent with two blanks. An Italian L1 interpreter translated the sentences including the idioms meaning into Chinese. A Chinese native speaker then reviewed the sentences to guarantee no correspondence between literal and figurative meaning in the two languages.

The initial letters of the words composing the target idioms were given, in order to avoid the possibility that participants may provide an alternative accurate response (Choi 2016, Peters 2012).

Example (5.5):

> 欢迎！您好！我给您介绍Matteo。 他帮助我做所有的事和解决实际问题:他是我的助理人员。
>
> Buongiorno e benvenuti. Vi presento Matteo: mi aiuta e si occupa della logistica. È il mio B______________ D_____________.
>
> *Good morning and welcome. Let me introduce you Matteo: he always helps me and cares about the logistic issues. He's my* R_________ *H* ___________ *man.*

This test is here meant to measure productive knowledge of both form and meaning. In the literature about collocational knowledge a different kind of test is usually employed in order to assess productive knowledge of form, i.e. subjects are presented with a node word and asked to provide the correct collocate (e.g. Webb t al 2013, Webb & Chang 2009). This kind of test was here avoided because it does not take into account the specific features of idioms. The target idioms are indeed collocations, but they are selected to be non-frequent. Therefore, providing more frequent collocates for the node words would be natural for the subjects. For this reason, a context is needed to highlight the requested meaning.

2) Form recognition: receptive knowledge of form

The test comprised 10 items, each consisting of four Italian sentences, which provided context for the target idiom. The four sentences were identical except for one of the words composing the idiom, and the participants had to choose the correct one. Distracters were selected in order to be semantically plausible, equally known for the participants and to have a similar frequency ranking (Long 2015).

Example (4.6):

a) Diego aiuta Chiara e lavorano sempre insieme. Diego è il piede destro di Chiara.

b) Diego aiuta Chiara e lavorano sempre insieme. Diego è il braccio destro di Chiara.

c) Diego aiuta Chiara e lavorano sempre insieme. Diego è la mano destra di Chiara.

d) Diego aiuta Chiara e lavorano sempre insieme. Diego è la gamba destra di Chiara.

a) Diego helps Chiara and they are always together: Diego is Chiara's right-foot man.

b) Diego helps Chiara and they are always together: Diego is Chiara's right-hand man.

c) Diego helps Chiara and they are always together: Diego is Chiara's right-arm man.

d) Diego helps Chiara and they are always together: Diego is Chiara's right-leg man.

The most common practice for assessing receptive knowledge of collocation form includes the use of Gyllstad's (2009) COLLMATCH and COLLEX tests. COLLMATCH uses a yes/no format, as participants are asked to decide whether the presented sequences occur frequently or not. As stated above, the target idioms were here selected not to be frequent, therefore such a test would not have been appropriate. COLLEX has a multiple-choice format: participants are provided with a node word and 3 collocates options, and they are asked to decide which one is correct. The test created for the present study replicates this principle, but provides an additional context, which is considered to be necessary given the non-literal meaning the idioms bear.

3) L2 to L1 translation: meaning recall, i.e. receptive knowledge of meaning

The test comprised 10 items, each composed of an Italian sentence including the target idiom in a partially-informative context and 3 possible Chinese translations. The participants were asked to choose the correct translation. The distracters were created in order to provide alternative but plausible figurative or literal meanings for the idioms.

Example (4.7):

Ieri ho fatto una gita fuori città. Ho toccato il cielo!

a) 昨天我爬了很高的山。

b) 昨天我为了短期旅行很高兴。

c) 昨天我坐了飞机。

Yesterday I took a trip out of town. I touched the sky!

a) I was very happy.

b) I went to a trip and climbed up a mountain.

c) I got on a plane.

Receptive translation is widely employed as a test for receptive knowledge of meaning (e.g. Laufer & Girsai 2008; Peters 2012; Szudarki & Carter 2016; Webb & Kagimoto 2011; Webb et al 2013). Multiple choices are usually not provided, but here such a format was necessary in order to avoid literal, valid translations of the word sequences.

4.5.6. Retrospective verbal reports – level of consciousness at the point of learning

One of the defining features of implicit knowledge is the lack of awareness at the moment of acquiring it. In other words, the subjects demonstrate an amount of knowledge but cannot describe it nor verbalize how and when it was created. A common procedure to verify whether subjects are aware of having learned language elements is to prompt them to verbalize patterns or items they have noticed while performing the experimental tasks (Gass & Mackey 2000; Rebuschat 2013). This procedure usually takes place after the treatment, i.e. during the debriefing session, and it is therefore defined as a retrospective verbal report. Empirical studies starting in the 70s (e.g. Broadbent, 1977; Broadbent & Aston, 1978; Broadbent, Fitzgerald, & Broadbent, 1986) used retrospective verbal reports in conjunction with the control-task paradigm, and demonstrated that task performance and verbalization draw on different cognitive systems, thus explaining why acquired knowledge might well be inaccessible to conscious recall. Debriefing interviews have been used with this this in mind in SLA research. For instance, Williams' (2005) subjects scored above chance in posttests about an artificial grammar they were exposed to in incidental-learning conditions, but were not able to explain and verbalize the rules underlying their answers in the test (for more details see § 1.2).

For the present study, the goal of the retrospective verbal reports is threefold: first, they provide the opportunity to verify the subjects' prior knowledge about the target idioms, in the absence of a pretest. Second, they verify whether the participants consciously noticed the enhancement devices, and if one enhancement format is more noticeable than the others. Third, they aim to investigate whether intentional learning took place.

In order to address such purposes, three questions were asked to the participants:

1) Had you ever encountered these idiomatic expressions before today's text?
2) Did you notice any enhancement device?
3) If so, did you pay more attention to the enhanced items? Why?

Retrospective verbal reports do contribute investigating such issues, but also present some methodological limitations that need to be taken into account. Berry and Dienes (1993) observed that low-confidence knowledge may not be verbalized even though it is conscious, and the same may happen for conscious rules with limited scope and validity (Dulany, Carlson & Dewey, 1984). Therefore, retrospective verbal reports can be insensitive and incomplete tools for measuring actual consciousness at the point of learning. The contribution of retrospective verbal reports becomes methodologically more rigorous and theoretically more relevant when their outcomes are triangulated with different and finer measures of attention like those provided by eye-tracking at the process level, which is performed in the current study.

4.6. Procedures

Before the main data collections took place, the experiment was piloted with 20 L1 Chinese students of Italian L2 enrolled in the Marco Polo – Turandot program at Accademia Lingua Italiana (Assisi, PG). The students' proficiency level was similar to the main data collections' subjects, as the A2 CILS tests showed. The same procedures were followed during the pilot and the main data collection, except for the way the aural input recording was played (see infra, this paragraph) and for the SPR presentation lists (see § 4.5).

In order to get a proper number of participants, three data collections took place (§ 4.2). The first two data collections were carried out during academic year 2017/2018, at Università degli Studi di Pavia and

Università degli Studi di Siena respectively. The experiment took place in the same time and venue students were used to attend their Italian classes at. Procedures were identical for the two first data collection, and included the following steps.

1) Proficiency test. Potential subjects carried out the A2-level CILS proficiency test a few days before the experiment.

2) Subjects with an inadequate level of proficiency were excluded. Scores from the proficiency test were used in order to pseudo-randomize the sample into 5 groups with no significant difference as of proficiency level.

3) The first information sheet and consent form were read and signed.

4) On the day of the experiment, the participants belonging to the experimental groups were divided into two rooms, according to the audio input they were supposed to be exposed to. One room gathered groups 1 and 4 (no aural enhancement), the other room gathered groups 2 and 3 (aural enhancement). During the pilot data collection, the choice had been made to provide each participant with the appropriate audio file and headphones. However, some of the subjects tried to stop the recording or listen to it more than once, which might have hampered the incidental nature of learning. Therefore, in the main data collections it was preferred to have a stricter control over the participants' exposure to the input and the recordings were played collectively. This is also the most common procedure in language classes, which adds ecological validity to the experimental treatment.

   Each student was given a hard copy of the reading-while-listening text, according to the written input he/she was supposed to be exposed to (with typographical enhancement for groups 1 and 3, without typographical enhancement for groups 2 and 4).

   Before starting the recording, participants were told some information about the genre of the story in the text and were made aware a general understanding test would have followed the reading. They were asked to pay attention to the meaning and to read the same words they heard in the recording, without stopping of reading ahead.

5) The recording was played in three phases, with a 5 minutes break between them. Chapters 1 and 2 (10 minutes), then chapters 3, 4 and 5 (13 minutes) and finally chapters 6 and 7 (9 minutes). The experimental treatment took a total of 45 minutes to be delivered.

6) After the treatment, participants were given a 15-minute coffee break, during which they had no access to the instructional material. Then the immediate posttest session began.

7) First, the general meaning comprehension test was delivered. No time constraint was given, but it took approximately 20 minutes to complete it. After the comprehension test, group 5 (control group) joined the experiment.

8) Second, participants individually took the self-paced reading test on a computer in a quiet room, with the researcher's supervision.

9) Third, the offline tests were delivered, in the following order: L1 to L2 translation, form recognition, L2 to L1 translation. Participants were given the hard copy of one test only after they had completed and handed over the preceding one. This procedure was necessary in order to prevent them from filling out the productive tests on the basis of the receptive tests input. No time constraint was imposed, but it took approximately 30 minutes to complete the offline tests.

10) Finally, on the same day the retrospective verbal reports were carried out.

11) Three weeks later, the delayed posttest session took place. The self-paced reading test and the three offline tests were repeated with the same procedures.

12) The second information sheet and consent form were read and signed.

The third data collection was carried out during academic year 2018/2019, at Università degli Studi di Pavia. Procedures for the third data collection were similar as of consent forms, proficiency testing, group pseudo-randomization and testing (steps 1-2-3-6-7-8-9-10-11-12). However, the two WM tests were delivered after the proficiency test. Moreover, the reading-while-listening text was not provided in hard copy. Instead, subjects read it on a computer screen while their eye movements were recorded. Therefore, the experimental treatment was delivered to one student at the time in the eye-tracking laboratory (for more details, see §4.5 section about eye-tracking). The text and the audio file were identical to the first and second data collection, therefore the treatment took approximately the same time to be delivered.

Chapter Five: Results

5.1. Introduction

The data collected according to the design described in the previous chapter were analyzed in order to answer the research questions:

1. Can formulaic sequences be learned incidentally through exposure to bimodal presentation of reading passages, without any explicit instruction?
    a. If knowledge of the target formulaic sequences is gained, is this knowledge implicit?
    b. If knowledge of the target formulaic sequences is gained, is this knowledge explicit?
2. Does adding enhancement to the first two occurrences of the target formulaic sequences affect learning, and if so, is one enhancement format among typographical, aural or both, more effective?
3. What is the level of consciousness at the point of learning in enhanced incidental learning conditions?
4. What is relationship between level of consciousness at the point of learning and the kind of knowledge gained?

As reported in the methodological section, three data collections during two subsequent academic years were carried out. The initial sample included a total of 104 students. However, 21 of them were discarded from the study due to (i) skipping the delayed posttest session, (ii) not reaching the appropriate level of proficiency, and/or (iii) scoring less than 75% on the comprehension test. Therefore, data from a total of 83 subjects were included in the analysis. Results from the posttests of the whole 83-subject sample are included as dependent variables in "Experiment 1" (§ 5.2).

In addition to the posttests, the subjects from the third data collection also had their eye movements recorded at the process level. Moreover, their working memory was measured. Therefore, for this 31-subject sample, a separate analysis was carried out. It is reported in the "Experiment 2" section (§ 5.3).

The last section in the chapter summarizes the major findings, which are dealt with in the following Discussion chapter. These significant results are also illustrated through graphs throughout the chapter.

All of the analyses were run with IBM SPSS statistics software, version 25.

5.2. Experiment 1

5.2.1. Sample, individual differences and comprehension test

The final sample for experiment 1 was composed as shown in Table 5.1, for a total of 83 experimental subjects.

Table 5.1. Experiment 1. Sample composition.

Group	**n**	**Treatment**
Group 1	18	Typographical enhancement (TE)
Group 2	17	Aural enhancement (AE)
Group 3	17	Typographical and aural enhancement (TAE)
Group 4	15	Increased frequency only (IFO)
Group 5 – control	16	No treatment

Proficiency is the only individual difference accounted for in experiment 1, and it was assessed by means of an A2 CILS exam. In each data collection, the sample was divided into experimental groups blocking on proficiency. When the data from the three data collections were merged into one sample, proficiency was checked and found not to be significantly different among the groups. Descriptive statistics are reported in Table 5.2.

Table 5.2. Experiment 1. Mean (SD) proficiency scores in each group (k=100)

	Group 1 – TE (n=18)	Group 2 – AE (n=17)	Group 3 – TAE (n=17)	Group 4 – IFO (n=15)	Group 5 – Control (n=16)
Proficiency	81.9 (8.02)	80.6 (11.1)	82.5 (8.1)	80.4 (8.9)	79.8 (11.1)

The results of the Kolmogorov–Smirnov and Shapiro–Wilk normality tests were significant; therefore, a non-parametric test was employed to check for differences among groups, namely, a Kruskal–Wallis test. No significant difference emerged (p=.858).

Subjects performed an announced comprehension test after the reading-while-listening treatment. Table 5.3 shows the average comprehension test scores in each group after subjects scoring less than 75% were excluded.

Table 5.3. Experiment 1. Mean (SD) comprehension test scores in each group (k=100)

	Group 1 – TE (n=18)	Group 2 – AE (n=17)	Group 3 – TAE (n=17)	Group 4 – IFO (n=15)
Comprehension	82.5 (8.2)	84.4 (9.3)	85.8 (8.3)	83 (7.9)

In this case also, Kolmogorov–Smirnov and Shapiro–Wilk normality tests showed an abnormal distribution; therefore, a non-parametric test was employed to check for differences among groups. The Kruskal–Wallis test displayed no significant differences ($p=.654$).

5.2.2. Online measures: self-paced reading test (RQs 1a and 2)

The self-paced reading (SPR) test outcomes were reaction times (RTs), which were measured using the Paradigm software. In the context of the present study and research questions 1a and 2, it is relevant to investigate whether RTs to violations were significantly longer than RTs to correct idioms. Such a difference would suggest violations are harder to process, thus hinting at implicit sensitivity to the collocational properties of the target items.

The experimental sets employed in the main data collection were the result of two failed validations and a third successful one, as described in the methodological section (§4.5) and in the next paragraphs.

The same statistical procedures were employed in all of these analyses. First, outliers were trimmed, discarding RTs shorter than 200ms and those deviating from the mean by more than 2.5 standard deviations. Moreover, RTs in trials where the subject provided a wrong answer to the T/F comprehension question were discarded from the analysis. Details about the percentage of excluded observations are reported in each section.

After the data trimming, Kolmogorov–Smirnov and Shapiro–Wilk normality tests were carried out. In all four analyses (three validations and the main data collection), the normality tests were significant, i.e. the data distributions were skewed. Log and square-root transformations were attempted but did not result in the data becoming normally distributed. Therefore, in order to determine whether differences between RTs to correct idioms and violations were significant, non-parametric tests were employed—namely the Mann–Whitney U test. Effect sizes for Mann–Whitney tests were calculated through the following equation (Rosenthal, 1991: 19): $r=Z/\sqrt{N}$

The analysis was carried out separately for each experimental group in order to assess the effectiveness of each treatment. The performance of the different groups was examined by comparing the effect sizes, in the case of significant differences.

The analysis was repeated for both immediate and delayed posttests.

5.2.2.1. Test validation: pilot study and native speakers

Before the main data collections, a pilot study with a small sample of 18 subjects was carried out. The participants were Chinese learners of L2 Italian enrolled in the same exchange program as subjects in the main sample, only at a different institution (Accademia Lingua Italiana, Assisi [PG], Italy). Subjects were randomly assigned into four groups; therefore, all of them were exposed to the treatment. For this reason, as well as because of the limited sample size, both descriptive and inferential statistics were calculated for each experimental group and for the whole sample.

The pilot SPR test contained each target idiom in both experimental conditions, i.e. in both the correct form and as a violation. The correct idioms and the violations appeared in different sentence contexts, in two counterbalanced sets (§ 4.5).

Data trimming resulted in the loss of 38 observations (5.2%). This first version of the SPR test proved not to be effective, as no significant difference was found between RTs to violations and to correct idioms, in either the single experimental groups or the whole sample. The analysis was carried out for both the target segments and the spillover regions, i.e. the segments following the target ones (indicated as "Correct idioms + 1" and "Violations + 1" in the tables). Descriptive statistics are reported in Table 5.4. Given the abnormal data distribution, Mann–Whitney tests were employed for inferential statistics (Table 5.5).

Table 5.4. Pilot data collection. Mean (SD) RTs (ms) to correct idioms and violations

	Correct idioms	Violations	Correct idioms +1	Violations +1
Group 1 – TE (n=4)	851 (300)	1159 (1153)	899 (498)	922 (550)
Group 2 – AE (n=5)	1405 (1718)	946 (383)	1082 (954)	897 (640)
Group 3 – TAE (n=5)	1679 (1534)	1829 (1654)	1089 (1087)	1295 (1463)

Group 4 – IFO (n=4)	1179 (823)	1187 (1047)	757 (446)	806 (387)
Whole sample (n=18)	1308 (1308)	1292 (1199)	971 (830)	993 (913)

Table 5.5. Pilot data collection. Difference between RTs to correct idioms and RTs to violations (Mann–Whitney test)

	p value (target segments)	p value (spillover segment)
Group 1 – TE (n=4)	.76	.1
Group 2 – AE (n=5)	.38	.18
Group 3 – TAE (n=5)	.57	.21
Group 4 – IFO (n=4)	.22	.28
Whole sample (n=18)	.65	.61

This result was ascribed to the different sentences the target idioms were presented in, as different contexts might have affected the reading times. Therefore, a new SPR test was created, where each subject was again exposed to both conditions (correct form and violation of each target idiom), but the sentences providing context were identical. A sample of eight native speakers (NSs) carried out the two counterbalanced sets created in order to validate the test. The NSs were university student volunteers. Three observations were discarded (0.9%) because of RTs being more than 2.5 standard deviations longer than the means, while all of the answers to the content questions were correct, as was expected with NSs. Again, no significant differences emerged between RTs to correct idioms and violations, in either the target segments (p=.98) or the spillover regions (p=.09). Descriptive statistics are reported in Table 5.6.

Table 5.6. NSs validation (1). Mean (SD) RTs (ms) to correct idioms and violations (n=8).

Correct idiom	Violation	Correct idiom+1	Violation+1
925 (554)	936 (574)	594 (322)	673 (299)

In order to explain this outcome, it was hypothesized that since participants read the same sentence twice (once for the correct idiom and once for the violation), the repetition effect could have interfered with the processing of violations. Such hypothesis was confirmed, as RTs to the second occurrence of each target item resulted in significantly shorter RTs than the first occurrence, regardless of the experimental condition (Table 5.7, p=.003).

Table 5.7. NSs validation (1). Mean (SD) RTs (ms) to first and second occurrences (n=8).

1st occurrence	2nd occurrence
1092 (532)	770 (548)

Taking the pilot and first validation outcomes into account, a third SPR test was created, which exposed subjects to only one experimental condition per target idiom. This test was validated with 16 NSs from a population similar to the L2 experimental subjects (university students). Descriptive statistics are reported in Table 5.8.

Table 5.8. NSs validation (2). Mean (SD) RTs (ms) to correct idioms and violations (n=8).

Correct idiom	Violation	Correct idiom+1	Violation+1
956 (508)	1402 (1110)	662 (351)	818 (744)

Data trimming resulted in the loss of eight observations (2.5%).

Given the abnormal distribution, a Mann–Whitney U test was employed to assess whether RTs to violations were significantly longer than RTs to correct idioms. The test showed a significant difference in the target item segments (r=-.22, p=.009) but not in the spillover regions (p=.08). This outcome suggests the SPR test was performing as intended. Segments following the target idioms and violations are accounted for in the non-native speaker (NNS) analysis, in order to control for possible behavioral differences between L1 and L2 speakers.

5.2.2.2. Main data collection

As a result of data trimming, a total of 303 (9.2%) observations were discarded, due to both outlying RTs and wrong responses.

As in the pilot and validation analysis, data from both the target and the spillover segments were investigated. No statistically significant difference emerged from the spillover segments. With regards of the target segments, in the immediate posttest only the typographical enhancement (TE) group showed significant sensitivity to violations ($p=.03$, $r=.18$). Three weeks later, no group exposed to enhancement retained any knowledge, while the increased-frequency only (IFO) group showed statistically significant difference between RTs to correct idioms and violations. In other words, subjects exposed to no additional enhancement of the first two occurrences of the target items showed no knowledge in the immediate posttest, while they did in the delayed posttest. Tables 5.9 to 5.12 report descriptive and inferential statistics for both the immediate and delayed posttests (Figures 5.1 and 5.2).

Table 5.9. Experiment 1. Mean (SD) RTs (ms) by experimental condition and by group, in the immediate SPR posttest.

	Correct idiom	Violation	Correct idiom+1	Violation +1
Group 1 – TE (n=18)	2092 (1308)	2430 (1384)	1093 (772)	912 (655)
Group 2 – AE (n=17)	2413 (1275)	2636 (1536)	1030 (655)	1115 (740)
Group 3 – TAE (n=17)	2341 (1214)	2525 (1301)	1139 (855)	1149 (822)
Group 4 – IFO (n=15)	2380 (1304)	2589 (1181)	843 (457)	1088 (784)
Group 5 – Control (n=16)	3103 (2097)	2829 (1370)	983 (727)	1077 (778)

Table 5.10. Experiment 1. Differences between RTs to correct idioms and RTs to violations in the five groups in the SPR immediate posttest (Mann–Whitney test)

	Target		Target +1	
	Effect size (r)	p value	Effect size (r)	p value
Group 1 – TE (n=18)	.18	.03		.26
Group 2 – AE (n=17)		.33		.42
Group 3 – TAE (n=17)		.078		.14
Group 4 – IFO (n=15)		.97		.26
Group 5 – Control (n=16)		.97		.26

Table 5.11. Experiment 1. Mean (SD) RTs (ms) by experimental condition and by group, in the delayed SPR posttest.

	Correct idiom	Violation	Correct idiom+1	Violation +1
Group 1 – TE (n=18)	1914 (1318)	2117 (1498)	1061 (725)	987 (593)
Group 2 – AE (n=17)	2026 (1492)	2205 (1433)	1002 (804)	963 (618)
Group 3 – TAE (n=17)	1777 (948)	2115 (1204)	1025 (662)	1141 (762)

Group 4 – IFO (n=15)	1687 (1059)	2057 (1314)	841 (479)	1032 (760)
Group 5 – Control (n=16)	2023 (1016)	2201 (1254)	1153 (763)	1142 (711)

Table 5.12. Experiment 1. Differences between RTs to correct idioms and RTs to violations in the five groups in the SPR delayed posttest (Mann–Whitney test)

	Target		Target +1	
	Effect size (r)	p value	Effect size (r)	p value
Group 1 – TE (n=18)		.24		.92
Group 2 – AE (n=17)		.43		.64
Group 3 – TAE (n=17)		.13		.42
Group 4 – IFO (n=15)	.17	.04		.38
Group 5 – Control (n=16)		.51		.88

Since non-parametric tests don't allow controlling for covariates, a separate correlation was run between RTs and proficiency. The aim was to verify whether individual differences might have significantly affected the SPR performance. A Kendall's test resulted in a significant but very weak correlation (T=-.056**).

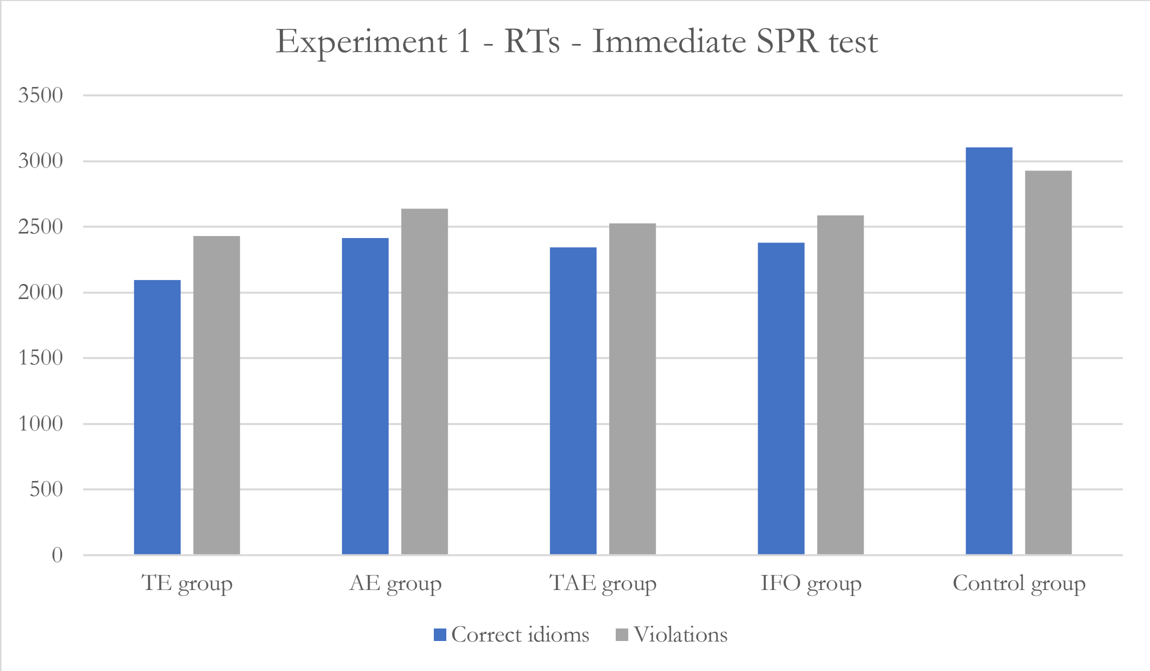


Figure 5.1. Experiment 1. SPR immediate posttest. Average reaction times (ms) to correct idioms and violations.

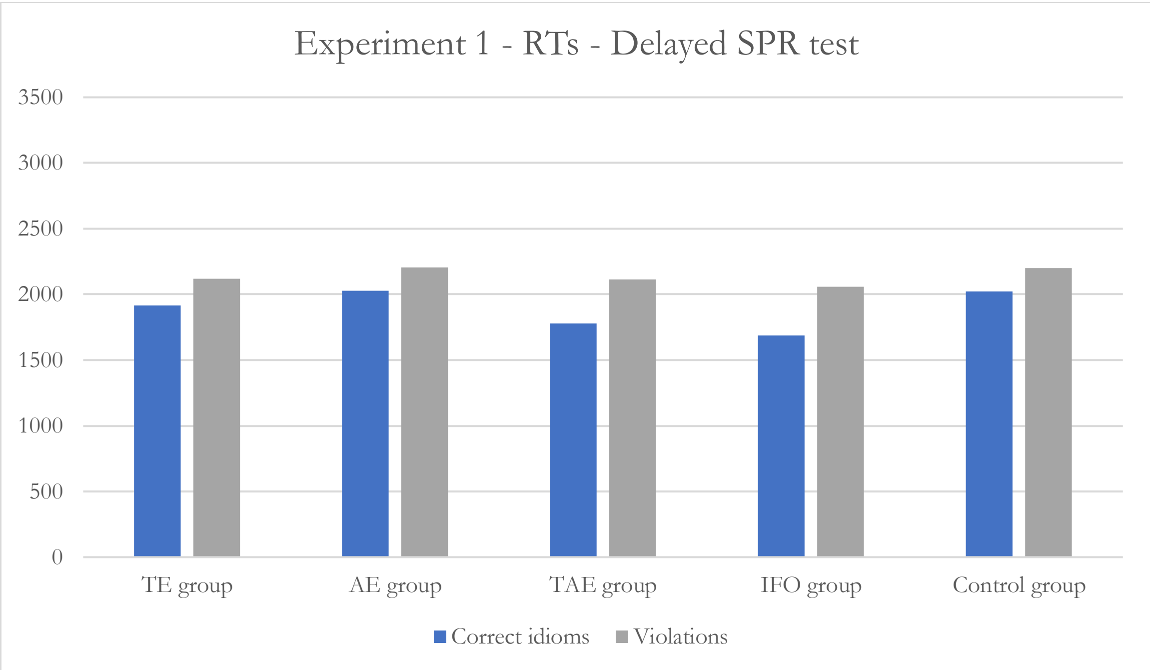


Figure 5.2. Experiment 1. SPR dealyed posttest. Average reaction times (ms) to correct idioms and violations

5.2.3. Offline measures (RQs 1b and 2)

A battery of three offline tests was created in order to assess explicit knowledge of the target items, which are the object of research questions 1b and 2. An L1-to-L2 translation test measured productive knowledge of form and meaning, a form-recognition test measured receptive knowledge of form, and an L2-to-L1 translation test was meant to measure receptive knowledge of meaning. The form-recognition and L2-to-L1 translation tests were in multiple-choice format; one point was assigned per correct answer. The L1-to-L2 translation test required the written production of the target items. One point was assigned per target item correctly produced, and minor orthographical errors were ignored.

The three tests were first carried out during the pilot study, in which they performed as expected. Therefore, no further validation was required.

The statistical analysis of the tests' scores aimed at comparing the experimental groups with the control group, in order to determine whether explicit knowledge was gained and retained. Moreover, the relative effectiveness of the different treatments was examined.

5.2.3.1. Pilot study

Nineteen subjects performed the offline tests as part of the pilot data collection. As previously mentioned, the pilot sample was divided into four groups. Therefore, all of the subjects were exposed to the treatment. Results of normality tests (Kolmogorov–Smirnov and Shapiro–Wilk) were non-significant, so a one-way ANOVA and Tukey post hoc tests were run among the four groups. No between-group difference reached statistical significance, as was expected due to the limited sample size. However, descriptive statistics (Table 5.13) showed a possibly stronger learning effect for subjects exposed to additional enhancement than for the increased-frequency only group.

Table 5.13. Pilot data collection. Offline tests scores: mean (SD), k=10

	L1-to-L2 translation	Form Recognition	L2-to-L1 translation
Group 1 – TE (n=4)	4.5 (4.7)	7.2 (2.7)	8.2 (2.3)
Group 2 – AE (n=5)	2.4 (3.5)	6 (2.8)	8.4 (1.3)

Group 3 – TAE (n=5)	4.4 (2.9)	7.4 (1.3)	8.2 (1.4)
Group 4 - IFO (n=5)	1.2 (1.1)	6.4 (2.6)	6.6 (1.5)

5.2.3.2. Main data collection

Descriptive statistics showed the L2-to-L1 translation test not to perform as expected due to a ceiling effect occurring also in the control group (Table 5.14). This can be attributed to a learning effect taking place during the SPR test and/or to ineffective distracters. As a consequence, L2-to-L1 translation scores were not taken into consideration in the inferential analysis.

Table 5.14. Experiment 1. Offline tests scores: mean (SD), k=10

	Immediate posttest			Delayed posttest		
	L1-to-L2 translation	Form Rec	L2-to-L1 translation	L1-to-L2 translation	Form Rec	L2-to-L1 translation
Group 1 – TE (n=18)	4.4 (3.2)	7.2 (2.4)	9.1 (1.1)	5.6 (3.1)	8.2 (1.7)	9.5 (0.7)
Group 2 – AE (n=17)	4.3 (3.3)	7.6 (2.5)	9.1 (1.1)	4.3 (2.8)	7.8 (1.9)	9.3 (0.6)
Group 3 – TAE (n=17)	4.3 (3.1)	7.6 (2)	8.7 (1.4)	4.8 (3.3)	7.6 (2.4)	9 (1.3)
Group 4 – IFO (n=15)	3.5 (3.2)	7.6 (2.1)	9.1 (1)	3.5 (2.8)	6.9 (2)	9.4 (0.8)
Group 5 – Control (n=16)	1.9 (2.8)	4.8 (2.4)	9 (1.1)	2.4 (2.2)	5.4 (2.7)	8.8 (1)

Results of normality tests for the main-data-collection scores were significant, even after log and square-root transformations. Therefore, non-parametric tests were employed to compare experimental groups. Specifically, the five groups were compared through a Kruskal–Wallis test, and Mann–Whitney tests were run as post hoc, applying a Bonferroni correction according to which p-values were considered significant at 0.01 (0.05/4=0.01). Since descriptive statistics showed no possible difference among experimental groups, post hoc tests were only run between each group and the control group, in order to limit power loss.

In the immediate posttest, the Kruskal–Wallis test showed a significant between-group difference in the L1-to-L2 translation test (H=13, p=.011) but not in the form-recognition test (H=8.7, p=.06). Mann–Whitney post hoc outcomes are reported in Table 5.15. Only the subjects exposed to typographical enhancement and typographical + aural enhancement (TE and TAE groups) showed a significant learning effect with a medium effect size at the productive level. When it came to the receptive level, the post hoc showed that all of the subjects exposed to the treatment significantly outperformed the control group in the immediate posttest, again with medium effect sizes (Figure 5.3).

Table 5.15. Experiment 1. Differences between experimental groups and the control group in the immediate offline posttests (Mann–Whitney test)

	Difference with groups 5 (control group, n=16) - L1-to-L2 translation	Difference with groups 5 (control group, n=16) - Form recognition
Group 1 – TE (n=18)	r= .43, p=.01	r= .41, p=.01
Group 2 – AE (n=17)	ns	r= .51, p=.003
Group 3 – TAE (n=17)	r=.49, p=.01	r= .53, p=.002
Group 4 – IFO (n=15)	ns	r= .52, p=.004

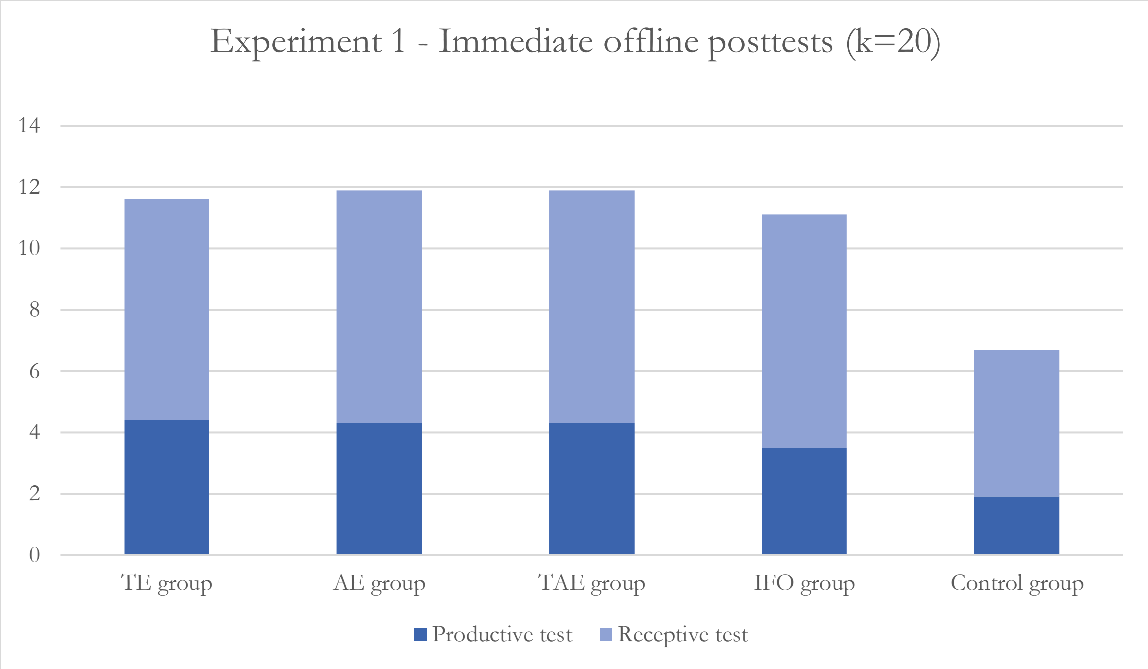


Figure 5.3. Experiment 1. Offline immediate posttests.

The delayed posttests showed a similar pattern. In the Kruskal–Wallis test, a significant between-group difference was found in both the L1-to-L2 translation test (H=11.7, p=.02) and the form-recognition test (H=10.4, p=.03). According to the Mann–Whitney post hoc tests, students exposed to typographical enhancement retained the productive knowledge after three weeks, again showing medium effect sizes. As for the receptive knowledge measured by the form-recognition test, TE, AE and TAE groups performed statistically significantly differently from the control group with medium effect sizes. The increased-frequency-only group instead showed no significant knowledge. In other words, only subjects exposed to enhancement, either typographical or aural, retained receptive knowledge of the target items after three weeks (Table 5.16, Figure 5.4).

Table 5.16. Experiment 1. Differences between experimental groups and the control group in the delayed offline posttests (Mann–Whitney test)

	Difference with groups 5 (control group, n=16) – L1-to-L2 translation	Difference with groups 5 (control group, n=16) – Form recognition
Group 1 – TE (n=18)	r= .48, p=.005	r= .51, p=.004

Group 2 – AE (n=17)	ns	r= .44, p=.01
Group 3 – TAE (n=17)	r=.41, p=.01	r= .41, p=.01
Group 4 – IFO (n=15)	ns	ns

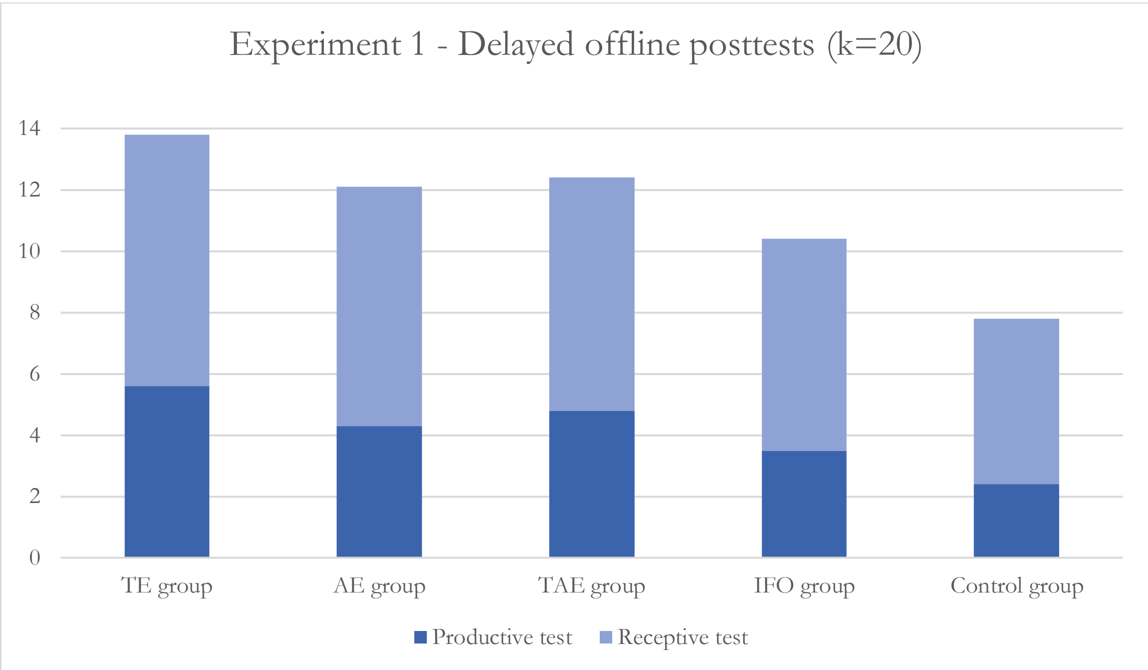


Figure 5.4. Experiment 1. Offline delayed posttests.

Proficiency was correlated with the offline tests' outcomes, and the correlation was significant and moderate (T= .44** for L1-to-L2 translation test; T= .36** for form-recognition test).

5.2.4. Retrospective verbal reports (RQ 3)

After the offline posttests, subjects carried out a stimulated recall. They were asked to answer three questions:

1) Had you ever encountered these idiomatic expressions before today's text?

2) Did you notice any enhancement device?

3) If so, did you pay more attention to the enhanced items? Why?

Crucially, the interview outcomes confirmed that the target items were not known to the participants before the treatment.

The main aim of retrospective verbal recalls was to investigate the level of awareness at the point of learning, though in an indirect manner, in order to address research question 3. Interviews outcomes showed typographical enhancement to be highly noticeable: 83% of the subjects in the TE group and 80% of those in TAE group claimed to have noticed the enhancement device during the treatment. On the other hand, no-one reported to be aware of the aural enhancement. Moreover, every subject reporting to be aware of enhancement declared to have intentionally paid attention to the target items, with some of the participants explicitly stating that the bolding led them to expect a vocabulary test about the enhanced items. Details about the retrospective recalls are reported in Table 5.17.

Table 5.17. Experiment 1. Retrospective verbal reports outcomes.

	Had already encountered the target items	**Noticed the enhancement device**	**Paid intentional attention to the target items**
Group 1 – TE (n=18)	0%	83%	83%
Group 2 – AE (n=17)	0%	0%	0%
Group 3 – TAE (n=17)	0%	80%	80%
Group 4 – IFO (n=15)	0%	-	-

5.3. Experiment 2

Experiment 2 followed exactly the same procedure as experiment 1, except it involved only a subsample of the subjects, who had their eye movements recorded at the process level, with the aim of addressing research questions 3 and 4. For the eye-tracking subsample, it was also possible to measure the subjects' working memory (WM). Table 5.18 reports the final sample composition.

Table 5.18. Experiment 2. Sample composition.

Group	n	Treatment
Group 1 – TE	7	Increased frequency + Typographical enhancement (TE)
Group 2 – AE	7	Increased frequency + Aural enhancement (AE)
Group 3 – TAE	6	Increased frequency + Typographical and aural enhancement (TAE)
Group 4 – IFO	4	Increased frequency only (IFO)
Group 5 – Control	7	No treatment

5.3.1. Individual differences and comprehension test

Both proficiency and WM were measured, and the descriptive statistics are reported in Table 5.19. Proficiency was assessed through an A2-level CILS exam. All of the subjects successfully passed the exam according to the certification scoring system.

WM was initially measured by means of two different tests, namely a digit span test and an operation span test. Both tests were piloted with a sample of European university students and were found to be valid. However, the digit span outcomes for the main-sample Chinese students showed a ceiling effect, which might be due to the strong memorization skills Chinese students develop in their education system. Therefore, only the operation span test was taken into account in the statistical analysis.

Table 5.19. Experiment 2. Individual differences: mean (SD) scores

	Proficiency	Working Memory
Group 1 – TE (n=7)	75.1 (6.3)	36 (4.2)
Group 2 – AE (n=7)	77.6 (14.4)	39 (9.8)
Group 3 – TAE (n=6)	81.1 (6.1)	42 (6.5)
Group 4 – IFO (n=4)	70.7 (4.4)	40 (10.8)
Group 5 – Control (n=7)	76.8 (4.4)	40.7 (3.2)

Individual differences were checked not to be significantly different among experimental groups using a one-way ANOVA (normally distributed data according to Kolmogorov–Smirnov and Shapiro–Wilk tests). No significant difference emerged for either proficiency (p=.42) or WM (p=.61).

The comprehension test the subjects performed after the treatment also acted as an inclusion criteria, as participants scoring lower than 75% were excluded from the study. Data from 10 subjects were discarded. After that, comprehension scores were checked for differences among groups, and a one-way ANOVA revealed no statistically significant differences (p=.33).

5.3.2. Eye-tracking measures: descriptive statistics

Early and late eye-tracking (ET) measures were collected for both the whole target idioms and the last words only. First pass reading time (FPRT), first pass fixation count (FPFC), total reading time (TRT), and total fixation count (TFC) were measured for the whole idioms. In addition to those, for the last words first fixation duration (FFD) and skipping rate (SK) were also analyzed. Since group 5 (control group) did not read the text, ET data were only collected for the four experimental groups. Descriptive statistics are reported in Tables 5.20 (whole idioms) and 5.21 (last words).

Two subjects were excluded due to calibration issues and invalid data. Reading times shorter than 50ms were discarded, and those deviating from the mean by more than 2,5 standard deviations were replaced with the cutoff value. Seventy-eight observations were discarded (2,3%), and 177 observations were modified (5,2%). Although the most common practice in the literature is to discard outliers, here the choice was made to replace the values instead, because of the already limited number of subjects and observations available (Field 2009).

Table 5.20. Experiment 2. Mean (SD) ET measures for whole idioms

		Mean (SD)
FPRT (ms)	Group 1 – TE (n=7)	500 (460)
	Group 2 – AE (n=7)	620 (568)
	Group 3 – TAE (n=6)	818 (601)
	Group 4 – IFO (n=4)	555 (466)
TRT (ms)	Group 1 – TE (n=7)	1125 (798)

	Group 2 – AE (n=7)	1192 (858)
	Group 3 – TAE (n=6)	1337 (767)
	Group 4 – IFO (n=4)	1082 (729)
FPFC	Group 1 – TE (n=7)	2.2 (1.8)
	Group 2 – AE (n=7)	2.3 (2.3)
	Group 3 – TAE (n=6)	3.1 (2.2)
	Group 4 – IFO (n=4)	1.6 (1.7)
TFC	Group 1 – TE (n=7)	5.1 (3.6)
	Group 2 – AE (n=7)	4.4 (3.8)
	Group 3 – TAE (n=6)	5.2 (3.4)
	Group 4 – IFO (n=4)	3.2 (3.1)

Table 5.21. Experiment 2. Mean (SD) ET measures for last words

		Mean (SD)
FPRT (ms)	Group 1 – TE (n=7)	252 (239)
	Group 2 – AE (n=7)	295 (268)
	Group 3 – TAE (n=6)	380 (292)
	Group 4 – IFO (n=4)	292 (271)
TRT (ms)	Group 1 – TE (n=7)	451 (436)
	Group 2 – AE (n=7)	492 (471)
	Group 3 – TAE (n=6)	557 (452)
	Group 4 – IFO (n=4)	459 (400)
FPFC	Group 1 – TE (n=7)	1.1 (1)
	Group 2 – AE (n=7)	1.07 (1.1)

	Group 3 – TAE (n=6)	1.4 (1.1)
	Group 4 – IFO (n=4)	.87 (.94)
TFC	Group 1 – TE (n=7)	2.1 (1.9)
	Group 2 – AE (n=7)	1.8 (1.9)
	Group 3 – TAE (n=6)	2.1 (2.1)
	Group 4 – IFO (n=4)	1.3 (1.5)
FFD (ms)	Group 1 – TE (n=7)	164 (138)
	Group 2 – AE (n=7)	188 (153)
	Group 3 – TAE (n=6)	242 (160)
	Group 4 – IFO (n=4)	196 (154)
SK (%)	Group 1 – TE (n=7)	17%
	Group 2 – AE (n=7)	13%
	Group 3 – TAE (n=6)	10%
	Group 4 – IFO (n=4)	13%

Kolmogorov–Smirnov and Shapiro–Wilk tests were employed to investigate the data distribution. No measure showed a normal distribution, despite log and square-root transformations. Therefore, non-parametric tests were employed for inferential statistics.

ET measures were correlated with both proficiency and WM scores by means of Kendall's test, as ET data were not normally distributed. The correlations were significant but very weak. It might be noted that correlations with WM, albeit weak, were stronger than those with proficiency (Table 5.22).

Table 5.22. Experiment 2. ET measures correlation with individual differences (Kendall's T).

	WM_OpSpan	Proficiency
FPRT	.19**	.08**
TRT	.16**	.06**

FPFC	.18**	.08**
TFC	.14**	.05**
FFD	.18**	.05**

5.3.3. Eye-tracking measures: between-group analysis (RQ 3)

Research question 3 addresses the levels of consciousness at the point of learning in the different learning conditions created, i.e. how the different enhancement formats affected attention. With the aim of dealing with this point, ET measures from the four groups were compared. Given the skewed distribution, non-parametric tests were employed—namely, a Kruskal–Wallis test to compare the four groups and then a series of six Mann–Whitney tests as post hoc tests. A Bonferroni correction was applied, so differences were considered significant when $p<.008$ ($.05/6=.008$). Effect sizes for Mann–Whitney tests were calculated with the following equation (Rosenthal, 1991: 19): $r=Z/\sqrt{N}$.

The analysis was repeated for both whole idioms and last words. First, data for all occurrences were analyzed together. Then data from the first two occurrences were analyzed separately, in order to observe more closely the enhancement effect on consciousness.

5.3.3.1. All occurrences, whole idioms

The Kruskal–Wallis test showed a significant difference for all of the ET measures (Table 5.23).

Table 5.23. Experiment 2. Between-group analysis for all occurrences and whole idioms.

	FPRT	TRT	FPFC	TFC
Kruskal–Wallis test	$H=67.2$ $p<.0001$	$H= 25.5$ $p<.0001$	$H=40.9$ $p<.0001$	$H=10.2$ $p=.016$

Therefore, post hoc Mann–Whitney tests were performed. Outcomes are reported in Table 5.24.

Table 5.24. Experiment 2. Post hoc analysis for all occurrences and whole idioms (Mann–Whitney test).

FPRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	r= .27, p<.0001	ns
Group 2 – AE (n=7)			r= .17, p<.0001	ns
Group 3 – TAE (n=6)				r=.2, p<.0001
Group 4 – IFO (n=4)				
TRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	r= .15, p<.0001	ns
Group 2 – AE (n=7)			r= .1, p=.003	ns
Group 3 – TAE (n=6)				r=.16, p<.0001
Group 4 – IFO (n=4)				
FPFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	r= .2, p<.0001	ns

Group 2 – AE (n=7)			r= .12, p<.0001	ns
Group 3 – TAE (n=6)				r=.18, p<.0001
Group 4 – IFO (n=4)				
TFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	ns	ns
Group 2 – AE (n=7)			ns	ns
Group 3 – TAE (n=6)				r=.13, p=.001
Group 4 – IFO (n=4)				

The inferential analysis showed that only Group 3 differed significantly from the other groups. This difference was significant but had small effect sizes, and it was larger in early than in late measures (Figure 5.5).

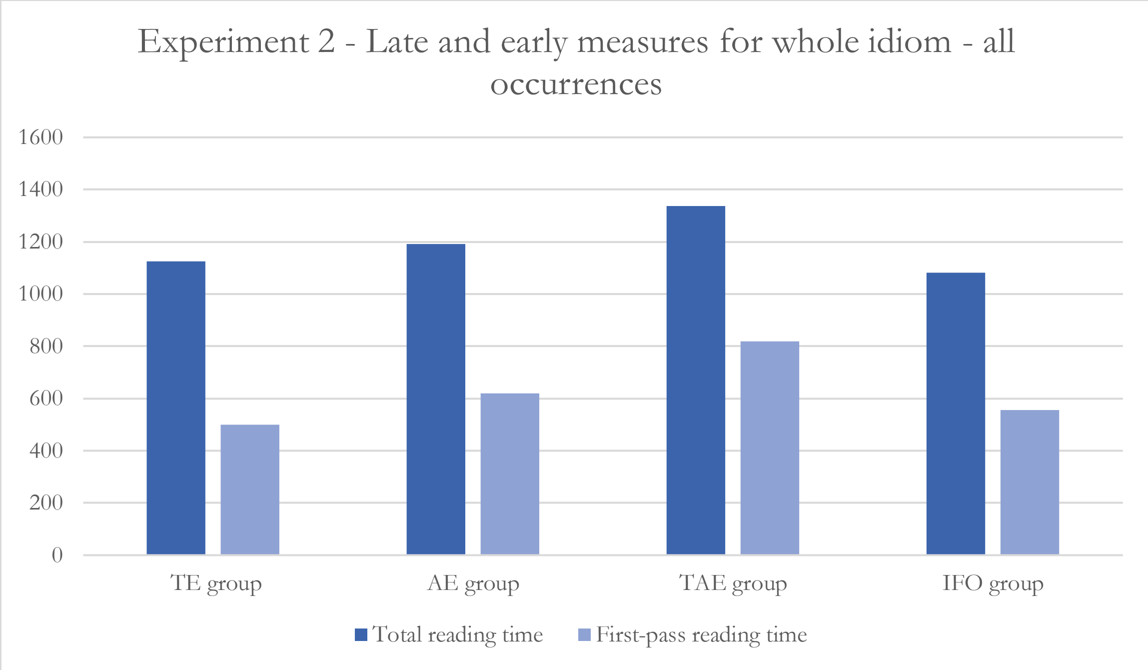


Figure 5.5. Experiment 2. Average TRT and FPRT (ms) for whole idioms - all occurrences.

5.3.3.2. All occurrences, last words

The Kruskal–Wallis test showed a significant difference in all of the ET measures except for total fixation count (Table 5.25).

Table 5.25. Experiment 2. Between-group analysis for all occurrences and last words.

	FPRT	TRT	FFD	FPFC	TFC
Kruskal–Wallis test	H=48.1 p<.0001	H=15.9 p=.001	H=60.6 p<.0001	H=18.3 p<.0001	ns

The last-word inferential analysis confirmed the pattern emerging from the whole-idiom data (Table 5.26). The TAE group is the only one reporting significantly more numerous and longer fixations, and this difference emerges with larger effect sizes in the early measures, especially the first fixation duration.

In other words, the between-group analysis carried out on data coming from all of the occurrences suggests that providing typographical-only or aural-only enhancement did not affect the subjects' behavior, as their ET measures were not significantly different from those of subjects who were exposed to increased frequency only. Only providing both typographical and aural enhancement combined seemed to significantly boost the participants' consciousness. However, it should be taken into account that both enhanced and non-enhanced occurrences were included in this analysis. Therefore, a learning effect taking place, as expected, in the final occurrences and thus speeding up those reading times might have affected the overall means and medians.

Table 5.26. Experiment 2. Post hoc analysis for all occurrences and last words (Mann–Whitney test)

FPRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	$r= .22$, $p<.0001$	ns
Group 2 – AE (n=7)			$r= .15$, $p<.0001$	ns
Group 3 – TAE (n=6)				$r=.15$, $p<.0001$
Group 4 – IFO (n=4)				
TRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	$r= .12$, $p<.0001$	ns
Group 2 – AE (n=7)			ns	ns
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				
FFD	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**

Group 1 – TE (n=7)		ns	r= .25, p<.0001	ns
Group 2 – AE (n=7)			r= .17, p<.0001	ns
Group 3 – TAE (n=6)				r=.14, p<.0001
Group 4 – IFO (n=4)				
FPFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	r=.13, p<.0001	ns
Group 2 – AE (n=7)			ns	ns
Group 3 – TAE (n=6)				r=.11, p=.003
Group 4 – IFO (n=4)				

5.3.3.3. Enhanced occurrences, whole idioms

In order to investigate more closely how being exposed to typographical or aural enhancement affected the subjects' level of attention, the between-group analysis was also run for the first two occurrences only, i.e. for the enhanced occurrences. Descriptive statistics are reported in Table 5.27.

Table 5.27. Experiment 2. Mean (SD) ET measures for enhanced occurrences and whole idioms.

		Mean (SD)
FPRT (ms)	Group 1 – TE (n=7)	536 (474)
	Group 2 – AE (n=7)	816 (652)

	Group 3 – TAE (n=6)	856 (699)
	Group 4 – IFO (n=4)	640 (526)
TRT (ms)	Group 1 – TE (n=7)	1321 (790)
	Group 2 – AE (n=7)	1616 (794)
	Group 3 – TAE (n=6)	1562 (895)
	Group 4 – IFO (n=4)	1041 (758)
FPFC	Group 1 – TE (n=7)	2.4 (2.2)
	Group 2 – AE (n=7)	3.4 (2.5)
	Group 3 – TAE (n=6)	3.2 (2.4)
	Group 4 – IFO (n=4)	2.6 (2.1)
TFC	Group 1 – TE (n=7)	6.2 (3.9)
	Group 2 – AE (n=7)	6.7 (3.2)
	Group 3 – TAE (n=6)	6.2 (4.3)
	Group 4 – IFO (n=4)	4.3 (2.7)

The Kruskal–Wallis test showed a significant difference for all of the ET measures (Table 5.28).

Table 5.28. Experiment 2. Between-group analysis for enhanced occurrences and whole idioms.

	FPRT	TRT	FPFC	TFC
Kruskal–Wallis test	H=14.3 p = .003	H = 28.8 p < .0001	H=9.3 p = .02	H=20.8 p < .0001

Therefore, post hoc Mann–Whitney tests were performed (Table 5.29). The enhanced-occurrences analysis for the whole idioms showed small effect size and significant differences between subjects exposed to enhancement and subjects in the increased-frequency-only group in the late-measures (total reading time and total fixation count).

Subjects exposed to typographical enhancement displayed shorter reading times and less fixation than expected, especially in the early measures.

Table 5.29. Experiment 2. Post hoc analysis for enhanced occurrences and whole idioms (Mann–Whitney test).

FPRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		r = .2, p=.001	r = .19, p=.002	ns
Group 2 – AE (n=7)			ns	ns
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				
TRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		r = .2, p=.001	r = .17, p=.006	ns
Group 2 – AE (n=7)			ns	r = .35, p<.0001
Group 3 – TAE (n=6)				r = .3, p<.0001
Group 4 – IFO (n=4)				
FPFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**

Group 1 – TE (n=7)		r = .17, p=.005	ns	ns
Group 2 – AE (n=7)			ns	ns
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				
TFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	ns	r = .23, p=.001
Group 2 – AE (n=7)			ns	r = .35, p<.0001
Group 3 – TAE (n=6)				r=.25, p=.001
Group 4 – IFO (n=4)				

5.3.3.4. Enhanced occurrences, last words

The same analyses were performed on the last word in the idioms. Descriptive statistics are reported in Table 5.30.

Table 5.30. Experiment 2. Mean (SD) ET measures for enhanced occurrences only and last words.

	Group	**Mean (SD)**
FPRT (ms)	Group 1 – TE (n=7)	256 (26)
	Group 2 – AE (n=7)	381 (277)
	Group 3 – TAE (n=6)	345 (304)

	Group 4 – IFO (n=4)	251 (269)
TRT (ms)	Group 1 – TE (n=7)	508 (452)
	Group 2 – AE (n=7)	679 (500)
	Group 3 – TAE (n=6)	598 (511)
	Group 4 – IFO (n=4)	412 (409)
FPFC	Group 1 – TE (n=7)	1.2 (0.92)
	Group 2 – AE (n=7)	1.6 (1.07)
	Group 3 – TAE (n=6)	1.3 (0.96)
	Group 4 – IFO (n=4)	1.08 (0.96)
TFC	Group 1 – TE (n=7)	2.5 (2.5)
	Group 2 – AE (n=7)	2.8 (1.9)
	Group 3 – TAE (n=6)	2.3 (2.2)
	Group 4 – IFO (n=4)	1.7 (1.6)
FFD (ms)	Group 1 – TE (n=7)	175 (128)
	Group 2 – AE (n=7)	222 (146)
	Group 3 – TAE (n=6)	221 (169)
	Group 4 – IFO (n=4)	160 (145)
SK (%)	Group 1 – TE (n=7)	10.7%
	Group 2 – AE (n=7)	5.8%
	Group 3 – TAE (n=6)	9.1%
	Group 4 – IFO (n=4)	21.6%

The Kruskal–Wallis test showed significant differences in all of the ET measures (Table 5.31).

Table 5.31. Experiment 2. Between-group analysis for enhanced occurrences and last words.

	FPRT	TRT	FFD	FPFC	TFC
Kruskal–Wallis test	H=17.7 p=.001	H=14.4 p=.002	H=11.5 p=.009	H=15.5 p=.001	H=14.8 p=.002

The post hoc analyses showed longer reading times and more numerous fixations in subjects exposed to aural enhancement. The differences in first-pass fixation counts between the aural enhancement group and all of the other treatments were significant, with small effect sizes (between r=.17 and r=.25). As for first-pass reading time, total reading time, and total fixation count, the only significant differences were between aural and typographical enhancement and between aural enhancement and increased frequency only, again with small effect sizes. For first fixation duration, no difference reached statistical significance according to the Bonferroni correction (Table 5.32, Figure 5.6).

Table 5.32. Table 5.32. Experiment 2. Post hoc analysis for enhanced occurrences and last words (Mann–Whitney test).

FPRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		r=.22, p<.0001	ns	ns
Group 2 – AE (n=7)			ns	r=.25, p=.001
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				
TRT	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		r= .17, p=.005	ns	ns

Group 2 – AE (n=7)			ns	r=.26, p<.0001
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				
FFD	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		ns	ns	ns
Group 2 – AE (n=7)			ns	ns
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				
FPFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		r=.19, p=.002	ns	ns
Group 2 – AE (n=7)			r=.17 p=.008	r=.25, p=.001
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				

TFC	**Group 1 – TE (n=7)**	**Group 2 – AE (n=7)**	**Group 3 – TAE (n=6)**	**Group 4 – IFO (n=4)**
Group 1 – TE (n=7)		r=.19, p=.002	ns	ns
Group 2 – AE (n=7)				r=.28, p<.0001
Group 3 – TAE (n=6)				ns
Group 4 – IFO (n=4)				

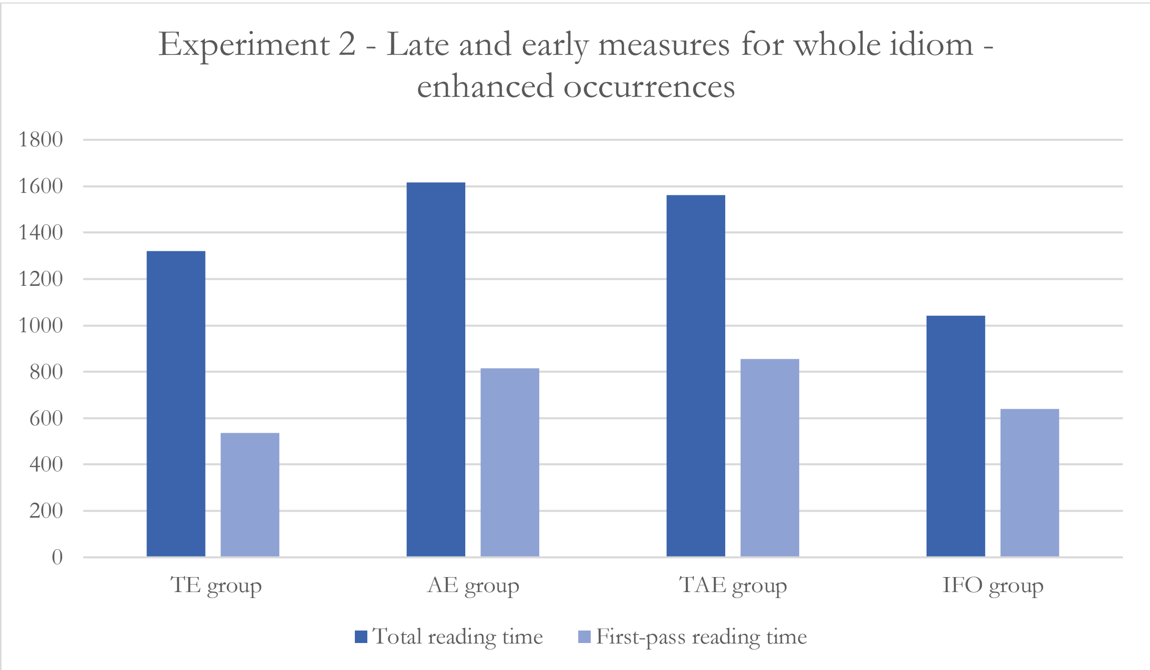


Figure 5.6. Experiment 2. Average TRT and FPRT (ms) for whole idioms - enhanced occurrences.

5.3.4. Eye-tracking measures: repetition effect (RQs 2 and 3)

Decreasing number and duration of fixations are usually regarded as indicators of growing familiarity with the target items and, therefore, signs of a learning effect having taken place. Eye-tracking is therefore capable of providing data not only about the levels of consciousness, but also about the learning process taking place. In other words, the analysis of repetition effect further contributes to addressing research question 2.

In order to investigate the learning process, Kruskal–Wallis tests were conducted comparing ET measures across the seven encounters in each group, for both the whole idioms and the last words. When the Kruskal–Wallis test highlighted a significant difference across encounters, Mann–Whitney post hoc tests were conducted, comparing the first encounter with each of the following and therefore applying a Bonferroni correction according to which p-values were considered significant at 0.008 (0.05/6=0.008). This way, it was possible to determine at what point the learning effect became apparent and whether its pattern was linear or not (Pellicer-Sánchez, 2016).

5.3.4.1. Whole idiom analysis

Table 5.33 reports descriptive statistics relative to each measure and each group across the seven encounters.

Table 5.33. Experiment 2. Mean (SD) ET measures for the experimental groups across encounters – whole idioms

FPRT (ms)	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	566 (474)	506 (476)	499 (423)	504 (489)	540 (493)	459 (487)	416 (370)
Group 2 – AE (n=7)	758 (626)	874 (678)	604 (563)	513 (506)	743 (563)	394 (427)	440 (414)
Group 3 – TAE (n=6)	1002 (697)	712 (677)	959 (559)	936 (605)	969 (516)	506 (476)	635 (470)
Group 4 – IFO (n=4)	634 (554)	645 (506)	552 (510)	496 (390)	569 (503)	572 (440)	409 (318)
TRT (ms)	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	1371 (691)	1272 (879)	1188 (832)	1060 (832)	1080 (761)	1078 (856)	828 (734)
Group 2 – AE (n=7)	1653 (742)	1580 (847)	1075 (820)	989 (859)	1003 (895)	832 (691)	935 (795)
Group 3 – TAE (n=6)	1802 (739)	1321 (976)	1437 (629)	1250 (629)	1256 (523)	1127 (813)	1139 (808)
Group 4 – IFO (n=4)	1019 (773)	1345 (755)	1045 (689)	1129 (728)	1009 (734)	1113 (768)	1197 (722)
FPFC	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	2.3 (1.6)	2.6 (2.4)	2.2 (1.7)	2.1 (1.7)	2.4 (1.8)	2.2 (2.1)	2 (1.7)

Group 2 – AE (n=7)	3.1 (2.3)	3.7 (2.8)	2.6 (2.3)	2.3 (2.4)	3.1 (2.3)	1.7 (1.7)	2.2 (1.9)
Group 3 – TAE (n=6)	3.7 (2.4)	2.7 (2.4)	3.4 (2.2)	3.5 (2.1)	3.6 (1.9)	2.1 (1.8)	2.6 (1.9)
Group 4 – IFO (n=4)	2.6 (2.2)	2.6 (1.9)	1.9 (1.4)	1.8 (1.3)	2 (1.5)	2.6 (1.8)	1.9 (1.2)
TFC	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	6.1 (3.4)	6.3 (4.4)	5.6 (3.7)	4.7 (3.5)	4.8 (3.5)	4.8 (3.3)	3.9 (3.03)
Group 2 – AE (n=7)	6.7 (3.2)	6.5 (3.1)	4.6 (3.4)	4.3 (3.8)	5.3 (3.7)	3.7 (3.03)	4.3 (3.5)
Group 3 – TAE (n=6)	7.3 (4.4)	5.2 (3.9)	5.2 (2.3)	4.7 (2.3)	4.7 (2.1)	4.7 (3.9)	4.6 (3.7)
Group 4 – IFO (n=4)	4.3 (2.8)	4.3 (2.6)	3.7 (2.6)	4.5 (3.6)	3.4 (2.5)	5 (3.1)	5.2 (3.06)

Subjects exposed to aural and typographical + aural enhancement showed a significant difference across encounters in all four measures, i.e. both early (FPRT, FPFC) and late (TRT, TFC) measures. In contrast, the typographical-enhancement group only displayed a significant learning effect for late measures. Finally, the increased-frequency-only group showed no increased familiarity with the target items across the seven encounters (Table 5.34).

Table 5.34. Experiment 2. Across-encounter analysis per group and ET measure – whole idioms (Kruskal–Wallis test)

	TE	AE	TAE	IFO
FPRT	ns	H=31,8, p<.0001	H=40,6, p<.0001	ns
TRT	H=25,1, p<.0001	H=54,1, p<.0001	H=39,8, p<.0001	ns
FPFC	ns	H=29,8, p<.0001	H=32,2, p<.0001	ns
TFC	H=22,3, p=.001	H=44,4, p<.0001	H=31,2, p<.0001	ns

Significant differences were further analyzed by means of Mann–Whitney post hoc tests (Table 5.35). Early measures took a larger number of encounters in order to display a significant decrease. Indeed, a significant decrease in length and number of first-pass fixations was apparent only at the sixth encounter. In contrast, late measures (total reading time, total fixation count) showed significant differences from the third or even second encounter on (Figures 5.7 and 5.8).

Table 5.35. Experiment 2. Post hoc tests outcomes for whole idioms measures across encounters (Mann–Whitney test).

FPRT	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	ns	ns	r= .31, p=.001	r= .25, p=.006
Group 3 – TAE (n=6)	ns	ns	ns	ns	r= .35, p<.0001	r= .26, p=.005
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns
TRT	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**

Group 1 – TE (n=7)	ns	ns	r= .23, p=.004	ns	ns	r=.4, p<.0001
Group 2 – AE (n=7)	ns	r= .37, p<.0001	r= .38, p<.0001	r= .29, p=.001	r= .5, p<.0001	r= .47, p<.0001
Group 3 – TAE (n=6)	r= .26, p=.003	r= .26, p=.003	r= .39, p<.0001	r= .41, p<.0001	r= .46, p<.0001	r= .45, p<.0001
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns
FPFC	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	ns	ns	r= .33, p<.0001	ns
Group 3 – TAE (n=6)	ns	ns	ns	ns	r= .33, p<.0001	ns
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns
TFC	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	r=.23, p=.005	r=.23, p=.006	ns	r=.32, p<.0001
Group 2 – AE (n=7)	ns	r= .32, p<.0001	r= .33, p<.0001	r= .27, p=.003	r= .44, p<.0001	r= .41, p<.0001

Group 3 – TAE (n=6)	ns	r= .32, p<.0001	r= .38, p<.0001	r= .4, p<.0001	r= .38, p<.0001	r= .41, p<.0001
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns

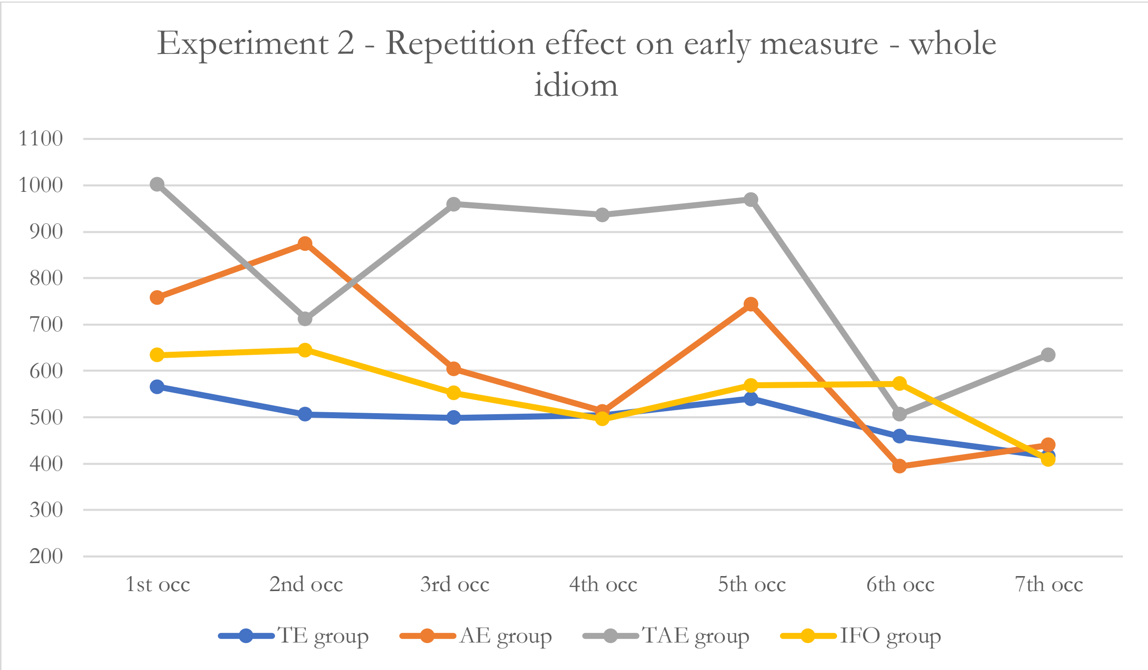


Figure 5.7. Experiment 2. Average FPRT (ms) per occurrence, per group - whole idiom.

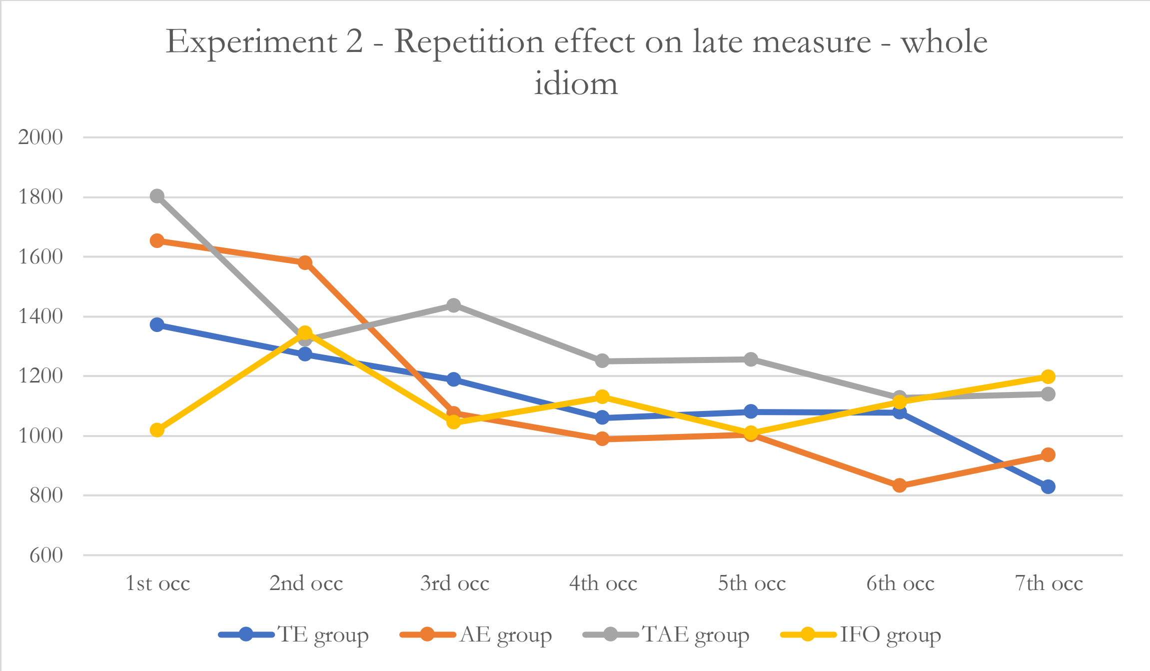


Figure 5.8. Experiment 2. Average TRT (ms) per occurrence, per group - whole idiom.

5.3.4.2. Last word analysis

As for the between-group analysis, all of the investigations were repeated for last-word measures. In this case, first fixation duration and skipping rates were taken into account as well. Data from skipping rates did not result straightforward.

Descriptive statistics are reported in Table 5.36.

Table 5.36. Experiment 2. Mean (SD) ET measures for the experimental groups across encounters – last words.

FPRT (ms)	1st occ	2nd occ	3rd occ	4th occ	5th occ	6th occ	7th occ
Group 1 – TE (n=7)	237 (213)	275 (218)	285 (258)	271 (290)	215 (266)	250 (219)	232 (196)
Group 2 – AE (n=7)	353 (237)	408 (310)	320 (275)	239 (246)	282 (295)	196 (222)	266 (228)
Group 3 – TAE (n=6)	327 (322)	319 (285)	470 (305)	434 (319)	376 (296)	334 (259)	356 (234)
Group 4 – IFO (n=4)	194 (240)	307 (287)	330 (273)	355 (320)	351 (310)	236 (213)	270 (219)
TRT (ms)	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	473 (447)	543 (458)	515 (473)	423 (430)	391 (454)	451 (440)	359 (323)
Group 2 – AE (n=7)	635 (458)	722 (539)	470 (465)	395 (430)	486 (528)	294 (326)	434 (404)
Group 3 – TAE (n=6)	635 (523)	561 (500)	641 (429)	533 (412)	501 (411)	477 (403)	554 (469)
Group 4 – IFO (n=4)	340 (397)	485 (415)	495 (473)	467 (414)	431 (396)	433 (330)	563 (364)
FPFC	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	1.1 (.8)	1.4 (.96)	1.2 (.94)	1.2 (1.2)	1.01 (1.2)	1.1 (.96)	1.01 (.76)
Group 2 – AE (n=7)	1.5 (.93)	1.8 (1.2)	1.3 (1)	1 (1.2)	1.2 (1.1)	.84 (.95)	1.2 (.93)

Group 3 – TAE (n=6)	1.4 (1.2)	1.2 (1.1)	1.7 (1.4)	1.5 (.95)	1.3 (.84)	1.2 (1.1)	1.4 (.81)
Group 4 – IFO (n=4)	.96 (.96)	1.2 (.96)	1.1 (.84)	1.2 (1.04)	1.1 (.99)	1.1 (.94)	1.3 (.75)
TFC	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	2.3 (2.7)	2.6 (2.2)	2.3 (2.3)	2.01 (1.9)	1.7 (1.9)	2.1 (2.1)	1.6 (1.4)
Group 2 – AE (n=7)	2.6 (1.9)	3.03 (2.04)	1.9 (1.9)	1.6 (1.8)	2.1 (2.3)	1.3 (1.3)	1.9 (1.6)
Group 3 – TAE (n=6)	2.6 (2.5)	2.1 (1.9)	2.3 (2)	1.9 (1.3)	1.8 (1.3)	1.8 (1.8)	2.2 (2.04)
Group 4 – IFO (n=4)	1.5 (1.4)	1.9 (1.7)	1.7 (1.8)	1.7 (1.6)	1.4 (1.2)	1.9 (1.4)	2.5 (1.5)
FFD (ms)	**1st occ**	**2nd occ**	**3rd occ**	**4th occ**	**5th occ**	**6th occ**	**7th occ**
Group 1 – TE (n=7)	176 (134)	174 (121)	184 (141)	148 (136)	131 (157)	162 (131)	175 (141)
Group 2 – AE (n=7)	226 (154)	217 (139)	198 (144)	162 (155)	188 (179)	145 (151)	176 (136)
Group 3 – TAE (n=6)	228 (158)	214 (179)	298 (149)	254 (164)	243 (159)	223 (171)	232 (130)
Group 4 – IFO (n=4)	118 (108)	203 (165)	241 (169)	201 (160)	244 (194)	172 (114)	194 (127)

SK	1st occ	2nd occ	3rd occ	4th occ	5th occ	6th occ	7th occ
Group 1 – TE (n=7)	17%	4%	17%	10%	31,8%	13%	8,8%
Group 2 – AE (n=7)	8.4%	3.3%	5%	8.3%	15%	15%	5.1%
Group 3 – TAE (n=6)	10%	8.3%	5%	10.3%	11.6%	6.6%	6.6%
Group 4 – IFO (n=4)	26%	16%	6%	16%	10%	10%	6%

As with the whole-idiom data, Kruskal–Wallis tests were conducted in order to analyze the five last-word measures in each group. Only subjects exposed to aural enhancement showed a significant repetition effect in all of the measures. The TE group showed no significant difference across repetitions in any of the measures, while the TAE group and the increased-frequency-only group displayed a slightly significant effect for repetition in the first-fixation duration data (Table 5.37).

Table 5.37. Experiment 2. Across-encounter analysis per group and ET measure – last words (Kruskal–Wallis test)

	Group 1 – TE (n=7)	Group 2 – AE (n=7)	Group 3 – TAE (n=6)	Group 4 – IFO (n=4)
FPRT	ns	$H=27.5$ $p<.0001$	ns	ns
TRT	ns	$H=31.3$, $p<.0001$	ns	ns
FFD	ns	$H= 14.4$, $p=.02$	$H=13.05$, $p=.04$	$H=13.1$, $p=.04$
FPFC	ns	$H=37.1$, $p<.0001$	ns	ns
TFC	ns	$H=34.6$, $p<.0001$	ns	ns

Significant differences resulting from the Kruskal–Wallis test were further investigated by means of Mann–Whitney post hoc tests (Table 5.38). The repetition effect in subjects exposed to aural enhancement did not display a linear pattern. Fixations of the third encounter were significantly shorter than those from the first encounter in both early and late measures. However, the fourth and fifth occurrences displayed an increase in fixation times, which decreased again at the sixth encounter. A similar pattern emerged form the number of fixations, as both first-pass and total fixation counts at the first encounter displayed a significant difference with the fourth and sixth occurrence, but not with the fifth and the seventh.

Table 5.38. Experiment 2. Post hoc tests outcomes for last-word measures across encounters (Mann–Whitney test).

FPRT	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	r=.26, p=.003	ns	r= .35, p<.0001	ns
Group 3 – TAE (n=6)	ns	ns	ns	ns	ns	ns
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns
TRT	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	r= .29, p=.001	ns	r= .37, p<.0001	ns
Group 3 – TAE (n=6)	ns	ns	ns	ns	ns	ns
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns

FFD	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	ns	ns	r= .27, p=.002	ns
Group 3 – TAE (n=6)	ns	r= .25, p=.005	ns	ns	ns	ns
Group 4 – IFO (n=4)	ns	r=.44, p=.002	ns	r= .34, p=.008	ns	ns
FPFC	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**
Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	r=.3, p=.001	ns	r= .38, p<.0001	ns
Group 3 – TAE (n=6)	ns	ns	ns	ns	ns	ns
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns
TFC	**1st vs. 2nd occurrence**	**1st vs. 3rd occurrence**	**1st vs. 4th occurrence**	**1st vs. 5th occurrence**	**1st vs. 6th occurrence**	**1st vs. 7th occurrence**

Group 1 – TE (n=7)	ns	ns	ns	ns	ns	ns
Group 2 – AE (n=7)	ns	ns	r= .3, p=.001	ns	r= .35, p<.0001	ns
Group 3 – TAE (n=6)	ns	ns	ns	ns	ns	ns
Group 4 – IFO (n=4)	ns	ns	ns	ns	ns	ns

5.3.5. Self-paced reading test (RQs 1a and 2)

After performing the treatment with their eye movements monitored, Experiment 2 subjects carried out the posttests like Experiment 1 participants did. As previously mentioned, RTs to correct idioms and violations were compared in each group in order to assess implicit sensitivity to the collocational properties of the target items, i.e. the object of research questions 1a and 2. Descriptive statistics are reported in Tables 5.39 and 5.40.

Table 5.39. Experiment 2. Mean (SD) RTs (ms), by experimental condition and by group in the immediate SPR test.

	Correct idiom	Violation	Correct idiom+1	Violation +1
Group 1 – TE (n=7)	2265 (1428)	2490 (1609)	1190 (745)	1136 (723)
Group 2 – AE (n=7)	2934 (1261)	2677 (1134)	1262 (831)	1160 (666)
Group 3 – TAE (n=6)	2335 (1203)	2244 (967)	1214 (891)	1460 (961)
Group 4 – IFO (n=4)	2844 (1328)	2852 (1627)	902 (465)	890 (594)
Group 5 – Control (n=7)	3111 (2595)	2865 (1453)	920 (695)	1225 (685)

Table 5.40. Experiment 2. Mean (SD) RTs (ms), by experimental condition and by group in the delayed SPR test.

	Correct idiom	Violation	Correct idiom+1	Violation +1
Group 1 – TE (n=7)	2205 (1235)	2308 (1613)	1148 (767)	1123 (694)
Group 2 – AE (n=7)	2474 (1350)	2810 (1309)	1014 (936)	1056 (728)
Group 3 – TAE (n=6)	2104 (854)	2345 (1207)	995 (600)	1233 (758)
Group 4 – IFO (n=4)	2109 (1982)	2264 (1768)	662 (237)	881 (558)
Group 5 – Control (n=7)	2292 (1039)	2536 (1639)	1148 (811)	1227 (684)

The data were abnormally distributed even after log and square-root transformations; therefore, a non-parametric test was employed. According to the Mann–Whitney tests, the difference in RTs between violations and correct idioms did not reach statistical significance in any of the groups (Tables 5.41 and 5.42). This can be attributed to the limited sample size and number of observations per subject (five per experimental condition). Taking into account these factors, ET data for repetition effect can be considered a more reliable measure of implicit knowledge in Experiment 2.

Table 5.41. Experiment 2. Differences between RTs to correct idioms and RTs to violations in the immediate SPR test (Mann–Whitney test)

	Target		Target +1	
	Effect size (r)	p value	Effect size (r)	p value
Group 1 – TE (n=7)		.53		.58
Group 2 – AE (n=7)		.46		.74
Group 3 – TAE (n=6)		.83		.45
Group 4 – IFO (n=4)		.71		.49
Group 5 – Control (n=7)		.55		.06

Table 5.42. Experiment 2. Differences between RTs to correct idioms and RTs to violations in the delayed SPR test (Mann–Whitney test)

	Target		Target +1	
	Effect size (r)	p value	Effect size (r)	p value
Group 1 – TE (n=7)		.74		.75
Group 2 – AE (n=7)		.43		.95
Group 3 – TAE (n=6)		.64		.19
Group 4 – IFO (n=4)		.79		.59
Group 5 – Control (n=7)		.51		.42

5.3.6. Offline measures (RQs 1b and 2)

Subjects in Experiment 2 carried out the battery of three offline tests, whose outcomes were analyzed and compared between groups in order to address research questions 1b and 2. As in Experiment 1, the distribution remained skewed despite log and square-root transformations, and non-parametric tests were therefore employed. The L2-to-L1 translation test did not perform as expected, as a ceiling effect was apparent also in the control group. As in Experiment 1, this was likely due to a learning effect taking place during the SPR test and/or to ineffective distracters. The L2-to-L1 translation test was therefore discarded and not included in the inferential analysis. Krukall–Wallis tests were employed to compare the five groups and Mann–Whitney tests were computed as post hoc tests, applying a Bonferroni correction, according to which p-values were considered significant at .005 (0.05/10=0.005). Table 5.43 reports descriptive statistics.

Table 5.43. Experiment 2. Offline tests scores: mean (SD), k=10.

	Immediate posttest			Delayed posttest		
	L1-to-L2 translation	Form Rec	L2-to-L1 translation	L1-to-L2 translation	Form Rec	L2-to-L1 translation
Group 1 – TE (n=7)	2.3 (1.9)	6.3 (1.9)	8.8 (1.2)	3.5 (2.1)	7.1 (1.8)	9.4 (.5)
Group 2 – AE (n=7)	2.3 (2.1)	6 (2.9)	9 (1.1)	3.1 (2.7)	7 (2.2)	9.1 (.6)
Group 3 – TAE (n=6)	1.7 (1.8)	5.5 (1.5)	8,1 (1.3)	3.4 (3.7)	5.8 (2.9)	8 (1.6)
Group 4 – IFO (n=4)	.5 (.62)	6 (2.5)	8.2 (.43)	1.6 (2.1)	6.5 (2.1)	9.2 (.4)
Group 5 – Control (n=7)	.5 (.71)	3.1 (1.1)	8.5 (1.3)	1.5 (1.4)	4.7 (3.1)	8.4 (1.05)

In the immediate posttest, the Kruskal–Wallis test showed statistically significant differences among the five groups in both the L1-to-L2 translation test ($H=78.9$, $p<.0001$) and the form-recognition test ($H=93.6$, $p<.0001$). According to the Mann–Whitney post hoc tests, groups exposed to enhancement significantly outperformed the control group in the translation test, while the increased-frequency-only group showed no learning effect at the productive level. However, all of the experimental groups, including subjects who read the text with no additional enhancement, displayed a learning effect at the receptive level, i.e. in the form-recognition test, with medium effect sizes. When it came to differences among experimental groups, descriptive statistics justified a closer investigation than the one run in Experiment 1. Therefore, despite the power loss implied, all of the groups were compared with one another. These tests showed that in the productive test, subjects exposed to enhancement significantly outperformed not only the control group, but also the increased-frequency-only group, with the analysis reporting medium effect sizes. In contrast, no significant difference emerged between subjects exposed to additional enhancement and the increased-frequency-only group at the receptive level (Table 5.44).

Table 5.44. . Experiment 2. Inferential statistics, immediate offline tests (Mann–Whitney test).

L1-to-L2 translation	**Group 1 - TE**	**Group 2 - AE**	**Group 3 - TAE**	**Group 4 - IFO**	**Group 5 – Control group**
Group 1 – TE (n=7)		ns	ns	r=.57, p<.0001	r=.57, p<.0001
Group 2 – AE (n=7)			ns	r=.48, p<.0001	r= .51, p<.0001
Group 3 – TAE (n=6)				r=.43, p<.0001	r=.44, p<.0001
Group 4 – IFO (n=4)					ns
Group 5 – Control (n=7)					

Form recognition	**Group 1 - TE**	**Group 2 - AE**	**Group 3 - TAE**	**Group 4 - IFO**	**Group 5 – Control group**
Group 1 – TE (n=7)		ns	ns	ns	r= .71, p<.0001
Group 2 – AE (n=7)			r=.24, p=.004	ns	r= .53, p<.0001
Group 3 – TAE (n=6)				ns	r=.68, p<.0001
Group 4 – IFO (n=4)					r=.58, p<.0001
Group 5 – Control (n=7)					

The delayed posttest data analyzed with the Kruskal–Wallis test confirmed the immediate-posttest pattern, showing a significant difference among groups in both the translation test (H=48.1, p<.0001) and the form-recognition test (H=29.1, p<.0001).

Table 5.45. Experiment 2. Inferential statistics, delayed offline tests

L1-to-L2 translation	**Group 1 - TE**	**Group 2 - AE**	**Group 3 - TAE**	**Group 4 - IFO**	**Group 5 – Control group**
Group 1 – TE (n=7)		ns	r=.28, p=.001	r= .45, p<.0001	r= .48, p<.0001
Group 2 – AE (n=7)			ns	r=.35, p<.0001	r= .35, p<.0001
Group 3 – TAE (n=6)				ns	ns
Group 4 – IFO (n=4)					ns
Group 5 – Control (n=7)					
Form recognition	**Group 1 - TE**	**Group 2 - AE**	**Group 3 - TAE**	**Group 4 - IFO**	**Group 5 – Control group**
Group 1 – TE (n=7)		ns	r=.25, p=.004	ns	r= .35, p<.0001
Group 2 – AE (n=7)			ns	ns	r=.39, p<.0001
Group 3 – TAE (n=6)				ns	ns
Group 4 – IFO (n=4)					r=.26, p=.005

Group 5 – Control (n=7)					

The translation test analysis showed that only the TE and AE groups' performances were significantly different from that of the control group. At the receptive level, the TE and AE groups showed a significant learning effect, as did the increased-frequency-only group. Finally, in both tests the TAE group scored significantly worse than the TE group (Table 5.45).

5.3.7. Retrospective verbal reports

The debriefing interview outcomes in Experiment 2 clearly confirmed the patterns that emerged from Experiment 1 about levels of consciousness at the point of learning in the different experimental conditions. No participant already knew the target items before the treatment, and no-one noticed the aural enhancement. As for typographical enhancement, 80% of the subjects noticed it, and therefore intentionally paid attention to the target items (Table 5.46).

Table 5.46. Experiment 2. Stimulated verbal reports outcomes

	Had already encountered the target items	**Noticed the enhancement device**	**Paid intentional attention to the target items**
Group 1 – TE (n=7)	0%	80%	80%
Group 2 – AE (n=7)	0%	0%	0%
Group 3 – TAE (n=6)	0%	80%	80%
Group 4 – IFO (n=4)	0%	-	-

5.3.8. Relationship between process and product level (RQ 4)

One of the aims of the current study is to investigate the relationship between the level of consciousness at the point of learning and the kind of knowledge gained, as expressed in research question 4. Therefore, outcomes from ET data, posttests, and retrospective verbal reports were analyzed together.

5.3.8.1. Eye-tracking measures and online posttest

The relationship between SPR outcomes and ET measures was investigated. With this aim, it was first necessary to assign to each subject a single value expressing if and to what degree they had gained implicit knowledge according to the SPR test. In order to do so, a Collocational Sensitivity Index (CSI) was calculated for each participant. The concept of Collocational Sensitivity Index was borrowed from Granena's (2013) Grammatical Sensitivity Index, which was calculated by subtracting RTs to grammatical items from RTs to ungrammatical items (GSI = ungrammatical RT – grammatical RT), for each subject. Similarly, CSI was here computed by subtracting RTs to correct idioms from RTs to violations (CSI = violations RT – correct idioms RT), repeating the operation for both immediate and delayed SPR test data. The only differences lay in fact that, in the present study, RTs were turned into z-scores before calculating the indexes, in order to limit the effects of individual differences in reading times and provide a better distribution to the data. The rationale underpinning the GSI and CSI is that RTs to violations should be longer than RTs to correct forms in subjects who have internalized the target structure. Therefore, the stronger the knowledge, the larger the difference between RTs to violations and correct items, i.e. the index value. In other words, subjects whose indices are positive are likely to have knowledge internalized about the target items, while participants with negative indices are not.

CSIs were correlated with the different ET measures, employing a non-parametric correlation test (Kendall's tau). Eye-movement measures did not have any significant correlation with CSI calculated on the basis of the immediate SPR test. Correlation between ET measures and delayed posttest CSI resulted in highly significant albeit weak negative correlations, for both whole-idiom and last-word data, with higher coefficients for early measures (Table 5.47).

Table 5.47. Experiment 2. Correlations between CSI and ET measures (Kendall's T)

Whole idioms – all occurrences	**CSI delayed SPR**
FPRT	-.097**
TRT	-.048**
FPFC	-.096**
TFC	ns
Last words – all occurrences	**CSI delayed SPR**
FPRT	-.058**
TRT	ns

FFD	-.065**
FPFC	-.058**
TFC	ns

This negative correlation may point to the fact that students who fixated less on the target items showed a stronger implicit sensitivity in the delayed SPR test. In order to investigate further, the analysis was repeated for the first two and the last two occurrences separately. The last two occurrences were selected because according to the repetition-effect analysis, the sixth encounter was the cut-off point where implicit knowledge was shown (§5.3.4). Interestingly, considering ET data from the first two occurrences, no significant correlation emerged. On the contrary, correlations between delayed CSIs and ET measures from the sixth and seventh occurrences were highly significant and showed higher coefficients than those emerging from the all-occurrences analysis, especially in the early measures (Table 5.48).

Table 5.48. Experiment 2. Correlations between CSI and ET measures (Kendall's T)

Whole idioms – 6th and 7th occurrences	**CSI delayed SPR**
FPRT	-.14**
TRT	-.084**
FPFC	-.15**
TFC	-.094
Last words – 6th and 7th occurrences	**CSI delayed SPR**
FPRT	-.11**
TRT	-.088**
FFD	-.08**
FPFC	-.14**
TFC	-.11**

This outcome may add validity to SPR and CSI as measures of implicit knowledge. However, it doesn't show any relationship between the number and duration of fixations and the implicit knowledge gained. In interpreting this data, however, the aforementioned limited number of observations available for the SPR test must be taken into account.

Besides the correlation, CSI data about the subjects were employed for a second kind of analysis. Subjects were divided into two groups: participants with positive CSI and participants with negative CSI. ET measures from these groups were then compared by means of a Mann–Whitney test. No significant difference emerged between groups formed according to the immediate CSIs. However, subjects with a positive delayed posttest CSI showed significantly shorter and fewer first-pass fixations for whole idioms than subjects with a negative delayed posttest CSI, albeit with very small effect sizes (FPRT: $r=.06$, $p=.01$; FPFC: $r=.05$, $p=.02$). In other words, this analysis might confirm the correlation outcomes, again showing a relationship between fewer and shorter fixations at the ET early-measure level and implicit sensitivity shown by the delayed SPR test.

5.3.8.2. Eye-tracking measures and offline posttests

The next step was to investigate the relationship between the subjects' reading behavior at the process level and their achievements regarding explicit knowledge. In order to do so, ET measures related to items that were learned (correct answers in the offline posttests) were compared to the ET measures related to items that were not learned (wrong answers in the offline posttests). The analysis was repeated for each of the two offline posttests (L1-to-L2 translation and form recognition) and for both immediate and delayed posttests (Table 5.49).

Table 5.49. Experiment 2. Differences between ET measures to learned vs. non-learned items (Mann–Whitney test).

Whole idioms	Immediate posttest		Delayed posttest	
	L1-to-L2 translation	**Form recognition**	**L1-to-L2 translation**	**Form recognition**
FPRT	r=.06, p=.01	ns	ns	ns
TRT	r=.07, p=.006	r=.07, p=.006	ns	ns
FPFC	r=.1, p<.0001	ns	ns	ns
TFC	r=.11, p<.0001	r= .07, p=.004	r= .1, p<.0001	r= .05, p=.04

Last words	Immediate posttest		Delayed posttest	
	L1-to-L2 translation	Form recognition	L1-to-L2 translation	Form recognition
FPRT	ns	ns	ns	ns
TRT	ns	r=.07, p=.003	ns	ns
FFD	ns	r= .05, p=.04	ns	ns
FPFC	r=.05, p=.05	ns	ns	ns
TFC	r=.06, p=.02	r= .08, p=.002	r= .06, p=.02	ns

Effect sizes are very small in this case, as well. However, some of the differences were significant. Namely, items that in the immediate posttests were determined to have been learned at the productive level had more numerous and longer fixations than non-learned items, with the differences being apparent in all of the examined ET measures for whole idioms and in fixation counts for last words. As for items learned at the receptive level, the difference reached statistical significance only for late measures in whole-idiom data, and in late measures and first fixation duration for last-word data. On the delayed posttests, learned items showed significant differences from non-learned items only in the case of total fixation count, in both tests for whole-idiom measures, and in the productive test only for last-word data.

5.3.8.3. Eye-tracking measures and retrospective verbal reports

Finally, reading behaviors were compared between subjects who reported having noticed the enhancement device and engaged in intentional learning and subjects who did not report any intentional learning. The difference was significant for both whole-idiom and last-word data in all of the ET measures examined, with larger effect sizes than those emerging from previous analyses (Table 5.50).

Table 5.50. Experiment 2. Differences in ET measures between subjects noticing and not noticing enhancement (Mann–Whitney test)

	Whole idioms	Last words
FPRT	r=.2, p<.0001	r=.2, p<.0001
TRT	r=.12, p<.0001	r=.1, p<.0001

FFD	-	r=.2, p<.0001
FPFC	r=.16, p<.0001	r=.1, p<.0001
TFC	r=.06, p=.01	r=.05, p=.04

5.4. Summary of major findings

5.4.1. Experiment 1

Experiment 1 aimed at comparing the relative effectiveness of different kinds of enhancement for the learning and acquisition of formulaic sequences.

A self-paced reading test was employed to assess implicit knowledge, while two offline tests measured the productive and receptive levels of explicit knowledge. Participants performed the tests immediately after the treatment and then again three weeks later.

The immediate self-paced reading test showed a significant learning effect only for subjects exposed to typographical enhancement. However, this knowledge was not retained according to the delayed test, which showed significant implicit sensitivity only for the increased-frequency-only group. RTs showed a weak correlation with the subjects' proficiency.

As for measures of explicit knowledge, in the immediate posttests, all of the experimental subjects showed a significant learning effect at the receptive level, while only groups exposed to typographical enhancement gained significant productive knowledge. Three weeks later, the productive-level pattern was confirmed, while at the receptive level, only subjects exposed to additional enhancement managed to retain the knowledge gained, with the increased-frequency-only group's scores showing no significant differences from the control group's. In this case, proficiency had significant, medium-effect correlation with test scores.

Finally, retrospective verbal reports highlighted that typographical enhancement (bolding, in this case) was detected by the subjects, thus triggering intentional learning.

5.4.2. Experiment 2

In addition to Experiment 1's goals, Experiment 2 also aimed at monitoring the participants' level of consciousness at the point of learning. Namely, the relationship between consciousness, the different kinds of enhancement, the repetition effect and the knowledge gained was investigated by means of eye-tracking.

First, the different treatments were compared. Considering all of the seven occurrences (of which only the first two were enhanced), the typographical-only and aural-only enhancement groups showed the same reading behavior as the increased-frequency-only group, while only subjects exposed to both typographical and aural enhancement had longer and more numerous fixations. In other words, only combining typographical and aural enhancement seemed to boost the subjects' attention, while adding aural-only and typographical-only enhancement didn't have any effect. However, when the analysis was repeated considering only the enhanced occurrences, the whole-idiom data showed that all of the three enhancement formats resulted in longer and more numerous fixations, especially in the late measures (total reading time and total fixation count).

The repetition effect confirmed an effect for the enhancement of the first two occurrences of the target items. Indeed, subjects exposed to increased-frequency-only showed no difference in their reading behavior from the first to the seventh occurrence of the target items, which suggests no learning effect took place. On the other hand, subjects exposed to additional enhancement displayed a decrease in number and length of fixations across the seven encounters, thus showing a growing familiarity and therefore knowledge being created. More specifically, the group who received typographical enhancement showed a learning effect only in the late measures. Groups exposed to aural enhancement and both typographical and aural enhancement showed a learning effect in both late measures (from the third exposure) and early measures (from the sixth exposure).

Posttest outcomes were analyzed as in experiment 1. No significant effect emerged from the self-paced reading test, probably due to the limited sample size. The offline tests partially confirmed experiment 1 results. In the immediate posttests, all of the experimental groups outperformed the control at the receptive level, while only the enhancement groups showed to have gained productive knowledge. Three weeks later, only the typographical enhancement and the aural enhancement groups retained productive and receptive knowledge.

The last step of the analysis was to investigate the relationship between level of consciousness at the point of learning and kind of knowledge gained, in other words, between ET and posttest data.

In order to analyze gained implicit knowledge, a collocational-sensitivity index was calculated for each subject from the SPR data. This index was correlated with the ET measures, and the correlation was weak but significant only for delayed posttest, with stronger effects on early measures and the sixth and seventh occurrences. The existence of a relationship between ET early measures and the online delayed posttest was confirmed by the test run to verify whether a significant difference existed as to reading behavior between subjects with a positive and a negative sensitivity index. The analysis showed that participants with positive indices (i.e. showing an implicit collocational sensitivity in the online delayed posttest) had fewer and shorter first-pass fixations at the process level.

When it came to explicit knowledge, items that the immediate posttest had shown to be learned at the productive level showed more numerous and longer fixations than items that were not learned. The same analysis at the receptive level showed differences in the reading behavior only for late measures.

Finally, the retrospective verbal reports outcomes were compared with reading behaviors. Subjects reporting to have detected the enhancement device and thus engaged in intentional learning indeed showed longer and more numerous fixations.

Chapter Six: Discussion

6.1. Introduction: study aims

The present study is aimed at investigating the effectiveness of enhanced incidental learning for the acquisition of L2 formulaic language.

Formulaic sequences are one of the most difficult items to acquire in a second language, due to a number of features, such as their ubiquity, dispersion, and the different levels of fixedness and compositionality they present. At the same time, formulaicity is a crucial part of mastering a second language, given its key pragmatic functions and the processing advantage it implies.

The teaching of formulaic language is highly problematic, as the poor performance of even advanced learners demonstrates. This difficulty can be ascribed to the interaction between pedagogic constraints, on the one hand, and formulaicity features, on the other. The pedagogic techniques reported in the literature range from the encouragement of fully intentional learning by means of explicit instruction to purely incidental learning, with no instructional intervention whatsoever. However, both these extremes have often proved ineffective.

Explicitly teaching formulaic language cannot constitute an effective option, due to the huge number of sequences that languages contain. Moreover, intentional learning results mainly in explicit knowledge, which is of little use during real-time communication.

Indeed, there is wide agreement about implicit knowledge being the desirable goal of language instruction, due to its automatic and durable nature and its availability for online language use. When it comes to formulaic language, the priority of implicit knowledge is even more apparent: collocational properties of vocabulary such as those regulating formulaicity are more likely to be learned and stored implicitly, since they involve statistical tallying of co-occurrences. At the same time, it must be considered that despite the ubiquity of formulaicity, most sequences occur very rarely in natural input. This makes purely implicit learning a process that, albeit possible, would require an amount of exposure and time that are not available in any language course.

The pedagogic technique proposed in the present work aims at taking into account all of the issues so far listed. The goal is to pursue enhanced incidental learning, i.e. to perform the least obtrusive intervention capable of affecting and speeding up learning while keeping it incidental. In order to do so, the Law of Contiguity is considered and exploited. According to the law, learning can involve two phases: first, a new item is consciously noticed, and this creates a memory trace in the learners' mind. Second, the existence of that memory trace triggers and makes possible unconscious, statistical learning and therefore the creation of implicit knowledge.

The pedagogic practice tested here aims at boosting both of these learning phases and does so through the manipulation of salience of the target items in a text. Firstly, the text is proposed in bimodal exposure conditions, which are known to favor incidental learning, comprehension and correct chunk segmentation. In order to favor the creation of a first memory trace of the target items, their salience is increased by applying input enhancement devices to their first two occurrences in the text. Then, with the aim of improving implicit, statistical learning, the extra enhancement is removed and the only intervention consists of increased frequency of the target items.

An additional consideration has to be made. The first learning phase illustrated above is meant to involve conscious noticing. However, noticing is regarded as resulting mainly in explicit knowledge, whereas as already stated, the present work aims at investigating a technique capable of resulting in implicit knowledge. Therefore, the hypothesis tested here is that triggering unconscious detection instead of conscious noticing in the first learning phase still results in significant learning gains. With this goal, the different enhancement devices tested here present growing levels of noticeability.

In summary, the research questions investigated were as follows:

1. Can formulaic sequences be learned incidentally through exposure to bimodal presentation of reading passages, without any explicit instruction?
    a. If knowledge of the target formulaic sequences is gained, is this knowledge implicit?
    b. If knowledge of the target formulaic sequences is gained, is this knowledge explicit?
2. Does adding enhancement to the first two occurrences of the target formulaic sequences affect learning and if so, is one enhancement format among typographical, aural or both more effective?
3. What is the level of consciousness at the point of learning in enhanced incidental learning conditions?
4. What is relationship between level of consciousness at the point of learning and the kind of knowledge gained?

To investigate these RQs, the following hypotheses were tested.

1. It is possible for L2 learners to gain implicit knowledge of formulaic sequences through bimodal exposure to a graded reader.
2. Enhancing the first two occurrences of the target items affects learning, and it can have different effects according to the type of enhancement employed.
3. When learners are exposed to enhanced incidental learning conditions, their level of awareness at the point of learning can be below the awareness threshold, depending on the noticeability of the enhancement devices employed.

4. There is a relationship between the level of consciousness at the point of learning and the kind of knowledge (implicit or explicit) gained.

A between-group design is used to address the RQs, with the experimental sample randomly divided into four experimental groups, each exposed to a different kind of enhancement device, plus a control group performing the post-tests only.

Dealing with RQs 3 and 4, i.e. with levels of consciousness, required use of eye-tracking technology. This was feasible only with a sub-sample of the experimental subjects, so the result, analysis and discussion are reported for two experiments. The first comprises the whole sample; the second concerns the eye-tracking subsample only.

6.2. Implicit knowledge gains: Research questions 1a and 2

The first part of research question 1 concerns the creation of implicit knowledge as a result of the EIL conditions created for the pedagogic treatment. With research question 2, the four experimental treatments involving enhancement devices with different degrees of obtrusiveness are compared.

In order to investigate these points, subjects performed a self-paced reading (SPR) test twice: the first time right after the treatment, and then again three weeks later. Self-paced reading tests measure the subjects' reaction time to single words or phrases when reading texts or sentences. In the current study, half of the experimental sentences which make up the test contained a violation of the target structure (i.e. a word in the idiom was replaced with a plausible synonym). What is relevant to measure in this case is the difference between reaction times to correct idioms and reaction times to violations. According to the anomaly-detection experimental paradigm, longer reaction times to violations than to correct idioms demonstrate that implicit sensitivity of the collocational properties of the target items has been internalized.

In experiment 2, eye-tracking (ET) observations about the repetition effect can be seen as an indication of knowledge creation.

6.2.1. Summary of results and discussion

Experiment 1. The results of the SPR test are less straightforward than expected. In the immediate post-test, descriptive statistics show that subjects from all four experimental groups needed more time to process violations than correct idioms, while for the control group the pattern was reversed. Despite this promising observation, the difference between reaction times to violations and correct idioms reached statistical significance only for the group exposed to typographical enhancement.

Outcomes from the delayed post-test further complicate the picture. The descriptive statistics present a similar pattern to the immediate post-test, with the exception of the control group, which in this case aligns with the experimental groups, probably benefitting from a test-related learning effect. When it comes to inferential statistics, the subjects exposed to typographical enhancement have apparently lost the sensitivity shown in the immediate post-test, while the increased-frequency-only group displays a significant difference between reaction times to violations and correct idioms in the predicted direction.

Delayed post-tests are considered more informative with regards to implicit knowledge, due to its greater durability. Taking this into account, it could be argued that in this case, the most effective treatment for the acquisition of implicit knowledge was the most implicit, i.e. the one involving increased frequency only, with no extra enhancement device added to the first occurrences of the target items. Such a result would be in line with what Toomer and Elgort (2019) found in their study of the relative effectiveness of reading only and reading with typographical enhancement. In their study, subjects exposed to the reading-only treatment showed implicit knowledge gains, whereas typographical enhancement resulted only in explicit knowledge. The proposed explanation for this is that typographical enhancement is too salient, thus triggering intentional learning of the bolded items and interfering with the tacit nature of the statistical process capable of leading to implicit knowledge creation.

Such lines of reasoning are indeed appealing and straightforward, but they cannot be fully applied in the context of the present experiment, because it does not account for two main outcomes: first, the presence of implicit knowledge for the typographical enhancement group in the immediate post-test, and second, the absence of knowledge in both post-tests for the aural enhancement group, which was exposed to a non-salient enhancement device. For these reasons, it must be noted that the SPR test outcomes were probably affected by methodological issues such as the limited sample size of the single groups. It is worth pointing out that the validation processes the SPR test went through with native speakers allows us to be reasonably sure that any issue regarding this test is not to be ascribed to its structure or items.

Experiment 2. Self-paced reading data from Experiment 2 was unsurprisingly not significant, since the single groups comprised only 4 to 7 subjects.

However, the analysis of repetition effects on eye-tracking data are considered capable of providing information about learning. In the literature, decreasing fixation duration and number or augmented skipping rates are regarded as signs of a growing familiarity with target words, i.e. an indication of learning. In order to detect this kind of evidence, eye-tracking measures from the first occurrence of the target items have been compared with each subsequent occurrence, for both the whole idiom and the last word only. Before reporting the results, it might be worth recalling the difference between early and late eye-tracking measures: early measures reflect automatic word recognition, while late measures witness more conscious and controlled processes.

In the whole-idiom data, subjects exposed to aural enhancement and aural + typographical enhancement showed a learning effect for both early and late measures. The typographical enhancement group displayed a decrease only on late measures, while increased frequency only did not result in any significant change in the reading behavior across the seven occurrences. Changes in early measures were detectable starting from the 6th occurrence, while late measures showed a significant decrease even at the 2nd or 3rd occurrence. This is coherent with the difference between early and late measure: since the former are a sign of automatization, they need a higher number of occurrences to be influenced. Significantly, early measures were not affected by the fact that the first two encounters were enhanced, while the following was not. On the contrary, the conscious-level processes late measures reflect can be affected by instance learning and therefore it is consistent for significant changes to be apparent early in learning processes.

When it comes to measures regarding the last word of the idioms, only the aural-enhancement group showed a significant decrease in number and duration of fixations, for both early and late measures.

In this case, aural enhancement seems to be the most effective form of instruction, since a growing familiarity with the target items can be observed in all measures, for both the whole idiom and the last word. The highly unobtrusive nature of aural enhancement is consistent with the fact that it was capable of affecting early measures, since it might be argued that learning remained incidental in this case, and therefore resulted in automatic knowledge. Typographical enhancement, on the other hand, is more noticeable and affected mainly late measures, which reflect conscious processes.

It needs to be noted that results from experiments 1 and 2 do not seem to point in the same direction when considering implicit knowledge gains. In Experiment 1, the IFO group seems to have the best performance. However, according to the eye-tracking measures of Experiment 2, that group was actually unaffected by the treatment. At the same time, the aural enhancement treatment resulted in a learning effect in Experiment 2, while no new knowledge is shown in Experiment 1's post-tests for the AE group.

Even though the sample for experiments 1 and 2 were different, such an outcome is problematic. The discrepancy may be attributed to self-paced reading being a coarser-grained measure than eye-tracking, and therefore failing to detect the knowledge created by subjects exposed to aural enhancement. On the other hand, the fact that eye-tracking measures show that attention is not affected by the IFO treatment might be read as a sign of unconscious learning which, with a three-week spacing effect, resulted in the implicit knowledge detected in the delayed SPR post-test.

For these hypotheses to be confirmed, more empirical data and investigation are needed.

Nonetheless, looking at the overall picture from both Experiments 1 and 2, it might be argued that typographical enhancement showed weaker results in relation to implicit knowledge than aural enhancement and increased frequency only. In Experiment 1, the TE group displayed gains only in the immediate post-test, which is regarded as less informative than the delayed post-test when it comes to implicit knowledge. In Experiment 2, typographical enhancement affected mainly late measures, which are related to conscious rather than unconscious processes.

In contrast, increased frequency only and aural enhancement showed effects, respectively, in the delayed post-test and in relation to early measures, both of which are considered to be closely related to automatized knowledge.

6.2.2. Implications

The data analysis does not result in a clear-cut indication of which treatment is the most effective for the creation of implicit knowledge. Therefore, no precise pedagogic direction can be provided without further research. However, with due caution, an interpretation of the available results seems to point to the fact that typographical enhancement is more likely to affect conscious learning and knowledge. On the other hand, treatments involving aural enhancement and increased frequency only may show results related to implicit learning and automatized knowledge.

If future research manages to confirm the direction the present results point to, the implication would be that typographical enhancement is noticeable enough to trigger intentional learning, and therefore it mostly results in the creation of explicit knowledge (see also § 6.3). Conversely, increased frequency only and aural enhancement go unnoticed and have a greater chance of eliciting incidental learning and implicit knowledge (see outcomes relative to RQ 3 in § 6.4 for details about noticeability of the enhancement devices).

As a consequence, an instructor aiming for implicit knowledge might want to avoid typographical enhancement, even if limited to the first two encounters with the target items, and instead, employ less obtrusive enhancement devices, like aural enhancement and increased frequency.

At a psycholinguistic level, this line of reasoning would imply that consciously noticing the first occurrence of unknown items hampers subsequent implicit learning. This would be in contrast with the Law of Contiguity, which maintains that noticing triggers statistical learning. A possible explanation for these contrasting outcomes may be that the Law of Contiguity refers to natural-input exposure, while in the present case, the noticed instances of the target items are followed by an unnaturally high number of further encounters in a relatively small amount of input. Such a concentrated learning condition is optimal for class instruction, but it greatly increases salience, and therefore requires typographical enhancement and noticing to be handled with more caution than when dealing with natural input.

6.3. Explicit knowledge gains: Research questions 1b and 2

To address the research questions about explicit knowledge gains, offline tests in pencil and paper format were administered, both immediately after the treatment and three weeks later. The aim of the tests was to assess four different levels of collocational knowledge, i.e. receptive and productive knowledge of both form and meaning.

Productive knowledge of form and meaning was measured by means of a first-to-second-language translation test, where subjects were asked to provide the target idioms in the context of translated sentences. Receptive knowledge of form was assessed through a multiple-choice test, requiring identifying the correct idiom among four likely options.

Finally, a second-to-first language translation test was employed to evaluate receptive knowledge of the meaning of the idioms. However, this last test turned out not to be a valid measurement tool, as a ceiling effect occurred in the control-group outcomes. This phenomenon can be explained in two main ways. First, the translation test had a multiple-choice format, and the distractors might have been too unlikely or easy to identify. Second, this test was the last to be administered, which means that at that point, subjects had already encountered the target idioms in context three times in the testing session (in the SPR test and in the two previous offline tests). Therefore, an assessment-related learning effect is a likely scenario, even though the pilot study did not reveal such a design issue. That said, there was no other option available for the testing of receptive knowledge of meaning. Moreover, this dimension of knowledge is the most likely to be created explicitly, with instance-learning processes similar to those taking place for single words. As a consequence, it was a reasonable choice to take the risk and compromise these data rather than those regarding implicit, productive or collocational knowledge. Results from the receptive test for knowledge of meaning are not considered in the discussion.

In Experiment 2, repetition effects on eye-tracking measures are analyzed as a way of assessing learning.

6.3.1. Summary of results and discussion

Experiment 1. The productive test proved to be more difficult than the receptive test, which is consistent with the previous literature (e.g. Laufer & Goldstein 2004; Laufer & McLean 2016).

Looking at research question 1b, the answer can be completely positive only at the immediate, receptive level, because only in the immediate multiple-choice post-test did all of the experimental groups score significantly higher than the control group. For the productive level and knowledge retention, distinctions need to be made among the different treatments, thus dealing with RQ 2.

Significant productive knowledge of meaning and form was created and retained only for subjects exposed to typographical enhancement and typographical + aural enhancement. The aural-enhancement and increased-frequency-only groups did not outperform the control group on the productive test. This result adds to those demonstrating what is occasionally the effectiveness of typographical enhancement for explicit knowledge gains (Boers et al 2017; Peters 2012; Sonbul & Schmitt 2013; Szudarski & Carter 2016; Toomer & Elgort 2019).

As mentioned, scores from the receptive test are generally higher, which mirrors the test's ability to show an initial stage of learning. At this level, in the immediate post-test, all of the subjects who were exposed to the reading-while-listening text showed a significant amount of collocational knowledge. However, the intervention consisting of increased-frequency-only did not result in a knowledge deep enough to be retained over time: only the groups exposed to additional enhancement of the first two occurrences (TE, AE and TAE) outperformed the control group three weeks after the treatment.

This outcome is consistent with the productive test results, as it confirms the effectiveness of typographical enhancement in improving learning. In addition, it shows that aural enhancement is capable of resulting in significant learning gains, as well, even though such gains are detected only at an initial stage of language acquisition, i.e. at the receptive level. This relative limitation on the effectiveness of aural enhancement is probably due to its unobtrusive nature, but the result is still compelling for different reasons. Firstly, few empirical data are available about aural enhancement, and to the best of my knowledge, no study has shown its effectiveness in the way the present does. Indeed, Cho and Reinders (2013) found no benefit for aural enhancement as compared to the reading-only treatment. Secondly, such an unobtrusive tool resulting in increased knowledge has implications for the noticing/detection debate, as will be discussed in section 6.5.

Experiment 2. Besides having their eye movements recorded at the process level, subjects from experiment 2 also took the offline post-tests.

The immediate post-tests mainly confirmed Experiment 1's patterns, with some differences. In the production test, subjects exposed to additional input enhancement of the first two occurrences of the target items (TE, AE and TAE groups) outperformed both the control group and the increased-frequency-only group. As in Experiment 1, in the receptive test, also the treatment involving only increased frequency resulted in significant learning gains.

The delayed post-test outcomes were less straightforward, mainly due to the poor performance of the group exposed to both typographical and aural enhancement. In fact, the TAE group is the only one not retaining knowledge from the immediate to the delayed offline post-tests and showing a significantly worse performance as compared to the other groups. Despite the fact that this result suggests that adding both typographical and aural enhancement can be detrimental for the retention of knowledge, two factors have to be taken into account in order to deal with these outcomes with due caution. Firstly, the data is a subsample of Experiment 1's, which means it is already accounted for in the inferential statistics previously reported and discussed, where no detrimental effect emerged for the TEA group. Secondly, the limited size of the sample for experiment 2 makes these results less trustworthy than those emerging from Experiment 1. Therefore, it is reasonable not to consider this poor performance of the TAE group as striking proof of the treatment's ineffectiveness.

The sample size issue is less relevant when it comes to eye-tracking data, because while each participant has one score per test in the offline assessment, with eye-tracking, each subject provides 70 data-points per measure (7 occurrences of 10 target items), for both the whole idiom and the last word only.

Among the eye-tracking data, those regarding the repetition effect are considered relevant for learning assessment, as explained for RQ 1a. The literature does not support a direct connection between late measures and explicit knowledge gains. However, late measures are known to be related to conscious processes, so it can be informative to take them into account when investigating RQ 1b and 2.

In the whole-idiom analysis, increased familiarity with the target items is shown in the late measures for subjects exposed to TE, AE and TAE. It is noteworthy that subjects who received aural enhancement (i.e. the AE and TAE groups) also showed effects for early measures, while for the TE group, the decrease in length and number of fixations was significant for late measures only. This may be related to the unobtrusive nature of aural enhancement, which is more likely to affect automatic processes.

When looking at the idiom's last-word-only data, the only group showing a significant repetition effect was the one exposed to aural enhancement, with increased familiarity displayed on both early and late measures.

Summing up, with regards to explicit knowledge gains, whole-idiom eye-tracking results confirm the benefits of enhancing the first two occurrences of the target items, as also emerged from Experiment 1's scores. Moreover, both whole-idiom and last-word-only data reinforce the finding from Experiment 1 about the effectiveness of aural enhancement.

6.3.2. Implications

On a pedagogic level, both Experiment 1 and Experiment 2 clearly point to the effectiveness of enhancing the first two occurrences of the target items in a reading-while-listening text, if the aim is improving explicit learning. This constitutes a precious indication for teachers, since adding enhancement to reading texts is an easy practice to implement, especially with typographical enhancement. Moreover, these processing conditions require the L2 learner to deal with texts, and therefore, to develop crucial skills, such as those related to reading comprehension, while at the same time boosting formulaic language learning.

Psycholinguistic relationships between the enhancement noticeability and the kind of knowledge gained are discussed in the section about RQ 4 (§ 6.5). However, it can be anticipated here that typographical enhancement (i.e. the most noticeable intervention) seems to result in the most robust effects on explicit knowledge gains. Therefore, a link between instruction noticeability and the of explicit knowledge might be a reasonable hypothesis.

Interestingly, the very interventions that looked more related to implicit knowledge gains (AE and IFO) were the least effective when it came to explicit learning, and vice versa: typographical enhancement proved minimally beneficial for implicit learning, whereas it significantly improved explicit knowledge. This finding, if confirmed by future research, can constitute further evidence for the claim that implicit and explicit knowledge are created through different processes and therefore require distinct pedagogic interventions for their development.

6.4. Levels of consciousness: Research question 3

One of the fundamental issues with the incidental learning literature is that once the conditions are set, scholars often fail to verify whether the learning process actually takes place incidentally, i.e. without intention. The third research question in the present study addresses this problem by measuring the levels of consciousness at the point of learning through retrospective verbal reports (Experiments 1 and 2) and eye-tracking (Experiment 2).

6.4.1. Summary of results and discussion

Experiment 1. In the context of Experiment 1, RQ 3 was addressed by means of retrospective verbal reports. Subjects were asked whether they had noticed the enhancement devices and whether they had intentionally tried to learn the target items.

Outcomes from the interviews clearly showed that typographical enhancement was consciously noticed by the subjects and that this led to intentional learning of the target items. Some of the participants even formulated the hypothesis that a test was going to be administered about the enhanced items.

On the other hand, none of the 34 participants exposed to aural enhancement detected it, nor reported having deliberately paid attention to the target items, i.e. having learned intentionally.

These outcomes imply that learning taking place in subjects assigned to the TE and TAE groups can hardly be considered incidental, at least with regards to the first two occurrences of the target items. In contrast, any significant difference in attention allocation between the baseline (the IFO group) and AE-exposed participants must be viewed as having resulted from unconscious processes.

Experiment 2. Experiment 2's patterns in the retrospective verbal report outcomes confirm those in experiment 1. Moreover, in this case, it was possible to validate and compare subjects' answers about their consciousness at the point of learning with their actual recorded reading behavior. This verification was implemented by comparing eye-tracking measures of subjects reporting no intentional learning with those of subjects claiming a conscious effort to learn the target items. This statistical analysis confirmed a significant difference between the two groups for all of the eye-tracking measures, thus validating retrospective verbal reports as a reliable instrument for the assessment of consciousness at the point of learning.

To measure the amount of attention paid to the target items in the different experimental conditions, early and late measures of number and duration of fixations were compared among the four groups. The analysis was repeated first considering all seven occurrences, and then the two enhanced occurrences only. Moreover, separated analyses were carried out for the whole idioms and for the last words only. Generally, whole-idiom and last-word analysis results converged, with data about whole idioms showing more definitive results. For this reason, the whole-idiom analysis in referred to here.

It should be noted that eye-tracking data analysis always consists of a relative rather than an absolute focus. In other words, experimental conditions have to be compared to a baseline for the significance of results to be determined. In this case, measures from the increased-frequency only group are taken as a baseline, since in that condition, no additional enhancement was added to the first two occurrences.

According to data from all seven occurrences (of which only the first two were enhanced), only subjects exposed to both typographical and aural enhancement (the TAE group) showed a significant difference from the IFO group. In contrast, subjects exposed to aural-only and typographical-only enhancement (the TE and AE groups) did not differ significantly in their reading behavior from participants reading the text with no additional enhancement device.

Looking at the enhanced occurrences only, the picture changes: all three enhancement groups (TE, AE and TAE) showed longer and more numerous fixations than the baseline IFO group.

These outcomes are consistent with existing findings, confirming the effectiveness of typographical enhancement in drawing attention to the target items (Cintrón-Valentín and Ellis 2015; Indrarathne and Kormos 2016; Issa et al 2015; Winke 2013). In addition, the present study offers several original findings.

Firstly, the effects of aural enhancement were investigated through eye-tracking, showing that adding a one-beat pause before and after the first two occurrences of the target items results in augmented attention to them. Crucially, combining this finding and the retrospective verbal report outcomes, it is possible to claim that such additional attention allocation took place without subjects being aware of it. Furthermore, adding aural to typographical enhancement also seems to affect reading behavior in the occurrences that were no longer enhanced.

Secondly, the combination of eye-tracking data and retrospective interviews outcomes showed that the augmented attention resulting from typographical enhancement is the result of a deliberate effort, i.e. from intentional learning, making defining such learning conditions as incidental questionable. Conversely, aural enhancement is capable of drawing attention to the target items without the subject being aware of it.

6.4.2. Implications

At a pedagogic level, data addressing RQ 3 provide practitioners with indications of the levels of consciousness each treatment involves. It is useful for teachers to know that adding typographical enhancement to target items in reading texts will probably result in intentional learning. Conversely, aural enhancement is unlikely to raise the level of consciousness above the awareness threshold.

Such information can be combined with that from the post-tests analyses, so that the instructor has the necessary tools to choose between pedagogic treatments aiming either at a faster, easier-to-measure, explicit knowledge gain or at triggering a slower process that is likely to result in implicit knowledge.

At a psycholinguistic level, the present findings imply that it is possible to raise the level of consciousness in a reader without them being aware of it, and that aural enhancement is a tool capable of such a result.

6.5. Relationship between consciousness and learning: Research question 4

Combining the process and product level outcomes as done when addressing RQs 1, 2 and 3 is sensible since the pedagogic treatments leading to ET and post-tests data were the same. Nevertheless, it has to be remembered that experiment 1 and experiment 2's samples were not the same. Therefore, for a rigorous, statistical investigation of the relationship between level of consciousness at the point of learning and the kind of knowledge gained, only data from experiment 2 were employed. In this way, the eye-tracking data describing the process level (level of consciousness) were only correlated with product-level outcomes (posttest reaction times and offline tests scores) from the very same subjects. Excluding experiment 1's post-test outcomes clearly resulted in a power loss for the statistical analysis, since experiment 2 had a smaller sample. However, that choice was made in order for the statistical analysis to be more rigorous and reliable.

To measure the relationship between levels of consciousness and implicit knowledge gains, eye-tracking measures were correlated with a collocational sensitivity index, and reading behaviors of subjects with a positive and a negative sensitivity index were compared. To address explicit knowledge gains, the number and duration of fixations of learned and not-learned items were compared.

6.5.1. Summary of results and discussion

Implicit knowledge. To run correlations and comparisons between eye-tracking and self-paced reading measures, it was necessary to associate each subject with a value describing his or her performance on the implicit knowledge post-tests. The collocational sensitivity index is such a value, since it is calculated from SPR reaction times to correct idioms and violations.

A weak but significant correlation between eye-tracking measures and collocational sensitivity scores emerged only for delayed post-test measures, and showed stronger effects for early measures and the sixth and seventh occurrences.

This relationship between self-paced reading delayed post-test data and reading behavior was confirmed by a different analysis: eye-tracking measures of subjects with a positive and a negative sensitivity index (i.e. with and without internalized knowledge of the target items) were compared. Subjects with a positive index on the delayed post-test showed significantly fewer and shorter first-pass fixations.

Simply, subjects who showed more implicit knowledge (positive and higher index on the SPR delayed post-test) also had reading behavior suggestive of a less conscious learning process (shorter and fewer fixations, especially at the early-measure level).

Explicit knowledge. Through use of the offline tests, it was possible to run the analysis by item rather than by subject. Therefore, eye-tracking measures of learned and non-learned items were compared. Idioms learned at the productive level showed significantly longer and more numerous fixations than non-learned items on both early and late measures. At the receptive level, the same pattern was shown only on late measures.

This means that if explicit knowledge of an idiom has been gained, it is likely that the idiom was the object of a conscious learning process, since subjects fixated significantly longer on it.

Implications

The analysis often resulted in small effect sizes, albeit highly significant, so more research and data are required to confirm claims implied by the current findings.

Nevertheless, these results confirm that unconscious learning is capable of resulting in the creation of knowledge, namely implicit knowledge. This is a relevant contribution to the noticing/detection debate, and it supports the role of detection in language acquisition, confirming previous findings, such as those reported by Williams (2005), Leung and Williams (2011; 2012) and Rebuschat and Williams (2012).

Furthermore, these findings prove that conditions for effective unconscious learning can be created in the classroom with relatively little effort.

Additionally, results from the offline tests show that explicit knowledge is the likely result of conscious learning. At the same time, intentional learning is less likely to result in implicit knowledge. For this reason, teachers involving their students in explicit teaching should not expect them to develop tacit, automatic knowledge of target items.

Despite the limited use of explicit knowledge in online communication, it can still be valuable for practitioners to have a pedagogic procedure capable of resulting in conscious learning and explicit knowledge gains while at the same time developing the subjects' reading and listening skills.

6.6. Limitations

The study has at least three limitations. First, the assessment of implicit knowledge gains proved problematic: results were not consistent between self-paced reading immediate and delayed post-tests, and between self-paced reading and eye-tracking outcomes. There are several possible causes. Self-paced reading is vulnerable to issues affecting reaction times, such as the need to press keys on the keyboard, i.e. an additional task that can be affected by individual differences. Moreover, in order to compare reaction times to correct idioms and to violations, each group had to be analyzed separately. Although the total sample size for this study is acceptable (83 participants), analyzing each of the five groups separately led to running statistics on limited samples of 15-18 subjects. Finally, for improved measurement of implicit knowledge, a second delayed post-test after a longer period of time (e.g. 2 months) would have been desirable.

A second issue with measurements regards eye-tracking. Empirical studies usually employ more powerful eye trackers than the one used in this study. The recommended sample rate for investigations related to words is 250Hz or more, which provides 250 XY coordinates of the gaze position per second. The study had to be conducted using an eye tracker with a 150Hz sample rate. Therefore, the measures might have lacked precision, leading to less accurate observation.

Finally, some choices had to be made to limit the number of variables and experimental groups. The decision was taken to not have frequency of encounters as an independent variable, which made the IFO group the baseline for eye-tracking analysis. However, increased frequency is already a form of enhancement, so it might have been informative to provide a baseline for comparisons consisting in a non-enhanced text or in matched non-idiomatic phrases.

6.7. Open questions and direction for future research

The main remaining open question concerns implicit knowledge gains. Self-paced reading outcomes did not allow a clear interpretation and did not match precisely the eye-tracking data. Therefore, more research is required in this area. Since self-paced reading did not prove to be effective enough for implicit knowledge assessment, future research might include a different kind of test, for instance, following a priming protocol or employing eye-tracking not only at the process level but also as a post-test. In addition, as mentioned in the limitation section, a second delayed post-test after a two- or three-month period would be desirable, in order to provide further evidence of the acquisition of implicit knowledge through enhanced incidental learning.

The second issue in need of further confirmation is the relationship between level of consciousness at the point of learning and kind of knowledge gained. The data reported in the present work points to a correlation between learning below the awareness threshold and implicit knowledge gains, on the one hand, and between more conscious learning processes and explicit knowledge improvements, on the other. Furthermore, pedagogic conditions promoting the acquisition of implicit knowledge seem not to be effective for explicit knowledge, and vice versa. Such claims provide a significant contribution to the noticing/detection debate, but they are supported by small effect sizes in the present study. Therefore, future empirical investigations focusing on this issue would be necessary.

It is crucial to point out that in order for this research area to be properly addressed, experimental designs need to include both online measures of awareness at the process level (e.g. eye-tracking or pupillometry) and assessments of implicit knowledge gains at the product level.

6.8. Final summary and conclusions

The present dissertation had two main aims. First, at a pedagogic level, the goal was to test the effectiveness of different kinds of enhanced incidental learning conditions for learning and acquisition of L2 idiomatic expressions. Second, from a psycholinguistic perspective, it addressed the relationship between level of consciousness at the point of learning and the knowledge gains.

The baseline treatment consisted of reading while listening to a graded reader where the frequency of the target items was artificially increased. Three kinds of additional enhancement with growing levels of noticeability were added to the first two occurrences of the target items, in order to test their effectiveness through a between-group experimental design.

Both explicit and implicit knowledge gains were measured, with offline post-tests and a self-paced reading test, respectively. In addition, eye-tracking was used to monitor the subjects' reading behavior at the process level, with the aim of assessing their level of consciousness during the pedagogic intervention.

Investigation of the pedagogic effectiveness of bi-modal enhanced incidental learning and of the different kinds of input enhancement resulted in the following findings, which answered research questions 1 and 2.

Typographical enhancement resulted in significant gains in explicit knowledge. Such gains were at both the productive and receptive level, and they were retained after a three-week period. The eye-tracking repetition-effect analysis confirmed that typographical enhancement affected conscious learning processes. A learning effect seemed to be apparent also at the implicit knowledge level, but it was lost by the time of the delayed post-test, so it cannot be considered as significant.

Adding both typographical and aural enhancement had the same effect as typographical-only enhancement on explicit knowledge gains, and no effect was detected on implicit knowledge.

Aural enhancement in the bi-modal EIL condition resulted in a significant and durable learning effect at the explicit, receptive knowledge level. In the self-paced reading test, no implicit-knowledge improvement was reported for aural enhancement, but eye-tracking data showed a growing familiarity with the target items in the early measures, i.e. those associated with automatic processes. This implies that a learning process might have been going on, even at too an early stage to be measured by the online post-test.

Exposing subjects only to increased frequency of the target items, with no additional enhancement devices, resulted in no increased familiarity with the target idioms across the seven occurrences of the pedagogic treatment, as assessed by the repetition-effect analysis of the eye-tracking data. At the explicit-knowledge level, increased frequency showed a limited effect as well: gains were measured only at the receptive level on the immediate post-test, but they were lost after three weeks. Unexpectedly, implicit knowledge gains were detected as a result of increased frequency only in the delayed self-paced reading test. This result was not confirmed in the immediate SPR post-test, nor in the eye-tracking data, so needs further empirical data.

In summary, the results about typographical enhancement confirm previous findings: it is useful if the aim is to improve explicit knowledge, but the evidence of its effects on implicit knowledge is questionable. In contrast, the results from aural enhancement are original to this study, and they help fill an important gap in the literature. Aural enhancement – and thereby, EIL – appeared capable of triggering the creation of explicit knowledge, even though at a lower degree than typographical enhancement. Additionally, and very importantly for unconscious detection and the development of implicit knowledge, eye-tracking data support the idea that implicit learning processes may also take place as a result of enhanced incidental learning.

When it comes to the psycholinguistic inquiry driving the present research, my findings provide an original contribution to the debate about the possibility of new knowledge being created below the awareness threshold (research question 4), as well as helping to determine which instructional conditions can really be considered incidental (research question 3).

Subjects claimed to have noticed the typographical enhancement, and to have engaged in intentional learning as a result. This was confirmed by their reading behavior. As a consequence, typographical enhancement, even if added to only two occurrences of the target items, cannot be considered an incidental-learning condition.

Data showed that explicit learning was the result of intentional learning, thus confirming the idea that explicit instruction is of little, if any, benefit for the creation of implicit knowledge.

In contrast with what is commonly reported about typographical enhancement, no subject noticed the one-beat pauses before and after the target items that constituted aural enhancement. Moreover, no subject reported having tried and learned the target idioms intentionally. Nevertheless, the reading behavior of participants exposed to aural enhancement did show an additional amount of attention devoted to the target items, as compared to the no-enhancement condition. These two findings combined imply that aural enhancement resulted in attention allocation taking place below the level of awareness. Therefore, aural enhancement can be considered a genuine example of an EIL condition.

Crucially, data suggests that the additional attention unconsciously paid to the idioms as a result of aural enhancement was capable of resulting in durable explicit knowledge and an increasing automatic familiarity with the target items throughout the reading text. This constitutes a critical contribution supporting the possibility that unconscious detection results in the creation of new knowledge.

A final analysis was carried out, considering the whole sample rather than the single treatments, in order to confirm the relationships between awareness and the kind of knowledge gained. Despite small effect sizes, the analysis linked lower levels of consciousness to implicit knowledge gains, and higher levels of consciousness to explicit knowledge gains. Again, this finding confirms the need for genuine incidental learning conditions in order to reach the primary goal of language instruction, i.e. the acquisition of implicit knowledge.

References

Anderson, J.R. (1982). Acquisition of a cognitive skill. *Psychological Review*, 89, 369- 406.

Arnon, I. & Christiansen, M. (2017). The Role of Multiword Building Blocks in Explaining L1–L2 Differences. *Topics in Cognitive Science,* 9(7), 621-636.

Arnon, I. & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. Special Issue: Parsimony and Redundancy in Usage-Based Models of Linguistic Knowledge. *Language and Speech,* 56(3), 257–264.

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language,* 62, 67–82.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Psychological Science*, 21(3), 170-176.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52, 323–363.

Beddeley, A. & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 8, 47-89.

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning-output instruction on the acquisition of the English past simple tense. *Language Teaching Research,* 9, 67–93.

Benati, A. (2013). The Input Processing Theory in second language acquisition. In M. del P. García Mayo, M. J. Gutierrez Mangado & M. Martínez Adrián (eds.): *Contemporary approaches to second language acquisition.* Amsterdam: John Benjamins, pp. 93-110.

Berry, D. C., & Dienes, Z. (1993). Implicit learning: Theoretical and empirical issues. Hove, UK: Erlbaum.

Bialystock, E. (1978). A theoretical model of second language learning. *Language Learning*, 28, 69-83

Bialystock, E. (1982). On the relationship between knowing and using linguistic forms. *Applied Linguistics*, 3, 181–206.

Bisson, M.-J., van Heuven, W.J.B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning,* 64(4), 855-877.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (eds.): *Grammar and second language teaching: a book of readings*. New York: Nwebury House, pp.19-33.

Boers, F. & Lindstromberg, S. (2012). Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics*, 83-110.

Boers, F., & Lindstromberg, S (2009). Optimizing a lexical approach to instructed second language acquisition. Basingstoke, UK: Palgrave Macmillan.

Boers, F., Demecheleer, M., He, L., Deconinck, H. & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, 27(2), 448-469.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheler, M. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. Language Teaching Research, 10(3), 245-261.

Bordag, D., Kirschenbaum, A., Tschirner, E. & Opitz, A. (2015). Incidental acquisition of new words during reading in L2: Inference of meaning and its integration in the L2 mental lexicon. *Bilingualism: Language and Cognition,* 18(3), 372-390.

Bowles, M (2003). The effects of textual enhancement on language learning. An online/offline study of fourth semester students. In P. Kempchinsky & C.E. Pineros (eds.): *Theory, Practice and Acquisition: Papers from the 6th Hispanic Linguistics Symposium and the 5th Conferencre on the Acquisition of Spanish and Portuguese*. Summerville, MA: Cascadilla Press, pp. 395-411.

Broadbent, D. (1977). Levels, hierarchies and the locus of control. *Quarterly Journal of Experimental Psychology,* 29, 181–201.

Broadbent, D., & Aston, B. (1978). Human control of a simulated economic system. Ergonomics, 21, 1035–1043.

Broadbent, D., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, 77, 33–50.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language,* 20, 136–163.

Bruton, A., García López, M., & Esquiliche Mesa, R. (2011). Incidental vocabulary learning: an impracticable term? *TESOL Quarterly*, 45(4), 759-768.

Bybee, J. L. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24, 215–221.

Carrol, G., Conklin, K. & Gyllstad, H. (2016). Found in translation. The Influence of the L1 on the Reading of Idioms in a L2. *Studies in Second Language Acquisition*, 38, 403-443.

Chang, A. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, 37(4), 652-663.

Chang, C.-S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System*, 52, 91–104.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31, 693–713.

Cho, M. & Reinders, H. (2013). The effects of aural input enhancement on L2 acquisition. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka, (eds.): *Noticing and second language acquisition: Studies in honor of Richard Schmidt*. Honolulu: University of Hawai'i, National Foreign Language Resource Center, pp. 133-148.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, 21, 403–426.

Christiansen, M. & Arnon, I. (2017). More Than Words: The Role of Multiword Sequences in Language Learning and Use. *Topics in Cognitive Science*, 1-10.

Cieslicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, 22(2), 115-144.

Cintrón-Valentín, M., & Ellis, N. (2015). Exploring the interface: Explicit focus-on-form instruction and learned attentional biases in L2 Latin. *Language Learning* 37(2), 197-235.

Cohen, A. & Robbins, M. (1976). Towards assessing interlanguage performance: the relationship between selecting errors, learner's characteristics, and learner's explanations. *Language Learning,* 26, 45-66.

Conklin, K. & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and non-native speakers? *Applied Linguistics*, 29, 72–89.

Conklin, K., Alotabi, S., Pellicer-Sánchez, A., Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. Second Language Research, 36(3), 257-276.

Conklin, K., Pellicer-Sánchez, A. & Carroll, G. (2018). Eye-Tracking: A Guide for Applied Linguistics Research. Cambridge: Cambridge University Press.

Day, R., Omura, C. & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language,* 7(2), 541-551.

Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (eds.): *Strategies in interlanguage communication*. London: Longman, pp. 175-195.

DeKeiser, R. M. (2015). Skill acquisition theory. In VanPatten, B. & Williams, J. (eds): *Theories in second language acquisition: an introduction*. Taylor and Francis, pp. 94-112.

DeKeyser, R. M. (2003). Implicit and explicit learning. In Doughty, C. J., & Long, M. H. (eds.): *Handbook of second language acquisition*. Oxford: Blackwell, pp. 313-48.

Dekeyser R. M. (1994). How implicit can adult second language learning be? In Hulstijn J. & Schmidt R. (eds.): *Consciousness in Second Language Learning*, AILA Review.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 875-87.

Doughty, C. (1991). Second language instruction does make a difference. Evidence from an Empirical Study of SL Relativization. Studies in Second Language Acquisition, 13, 431-469.

Dulany, D. E., Carlson, R. A., &Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, 113, 541-555.

Dulany, D. E., Carlson, RA., &Dewey, G. I. (1985). On consciousness in syntactic learning and judgment: A reply to Reber, Allen and Regan. *Journal of Experimental Psychology: General*, 114, 25-32.

Durrant, P. & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2), 125-155.

Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, 47(2), 157-177.

Durrant, P. & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research* 26(2), 163-188.

Ellis, A. W. & Young, A. W. (1988). *Human Cognitive Neuropsychology*. Psychology Press Ldt.

Ellis, N. (1994). Implicit and explicit learning – An overview. In N. Ellis (ed.): Implicit and explicit learning of languages. New York, NY: Academic Press, pp. 1.31.

Ellis, N. (1994). Implicit and explicit learning of languages. New York, NY: Academic Press

Ellis, N. (2001). Constructions, chunking, and connectionism: The emergence of second language structure. In Doughty, C. & Long, M. (eds.): Handbook of Second Language Acquisition. Oxford: Blackwell, pp. 2-66.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition,* 24(1), 143-188.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition,* 27, 305–352.

Ellis, N. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (eds): *Encyclopedia of Language and Education, 2nd Edition, Volume 6: Knowledge about Language*, pp. 1–13.

Ellis, N. (2018). Salience in usage-based SLA. In S. Gass, P. Spinner & J. Behney (2018). Salience in second language acquisition. New York, NY: Routledge, pp. 21-40.

Ellis R., Loewen S., Elder C., Erlam R., Philp J., Reinders H. (2009). Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching. Bristol: Multilingual Matters.

Ellis, R. (2004). The Definition and Measurement of L2 Explicit Knowledge. Language Learning, 54(2), 227-275.

Engle, R. (2002). Working Memory capacity as executive attention. Current Directions in Psychological Science, 11(1), 19-23.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 972–992.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29–62.

Faretta-Stutenberg, M. & Morgan-Short, K. (2011). Learning without awareness reconsidered: a replication of Williams (2005). In Granena G. et al. (eds): *Selected proceedings of Second Language Research Forum*. Somerville, MA: Cascadilla Proceedings Project, pp. 18-28.

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). London: Sage.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory and Cognition*, 43, 226–236.

Gass, S. (1988). Integrating research areas: a framework for second language studies. *Applied Linguistics*, 9(2), 198-217.

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21, 319–333.

Gass, S. M., & Mackey, A. (2000). Simulated recall methodology in second language research. Mahwah, NJ: Erlbaum.

Gass, S., Spinner, P. & Behney, J. (2018). Salience in second language acquisition. New York, NY: Routledge.

Giora, R. (2003). On our mind: salience, context and figurative language. Oxford: Oxford University Press.

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition,* 38(2), 177-215.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35, 483–517.

Goldshneider, J. & DeKeyser, R. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants. *Language Learning,* 51(1), 1-50.

Goo, J., Granena, G., Novella, M., & Yilmaz, Y. (2009). Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated. Paper presented at the Second Language Research Forum, Michigan State University.

Granena, G. (2013). Individual Differences in Sequence Learning Ability and Second Language Acquisition in Early Childhood and Adulthood. Language Learning, 63(4), 665-703.

Granena, G. (2016). Age differences, maturational constraints, and implicit and explicit L2 learning. *Brill Research Perspectives on Multilingualism and Second Language Acquisition*, 1(1), 9–44.

Granena, G. & Long, M. H. (2013). Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment. Amsterdam: John Benjamins.

Gries, S. & Ellis, N. (2015). Statistical Measures for Usage-Based Linguistics. *Language Learning*, 65(1), 228-255.

Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In A. Barfield & H. Gyllstad (eds.): *Researching collocations in another language*. New York, NY: Palgrave Macmillan, pp. 153-170.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition,* 32(3) 465-91.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237-258.

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, 31(6), 785–800.

Horst, M., Cobb, T. & Meara, P. (1998). Beyond a Clockwork Orange: acquiring secondo language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.

Houck, N., Robertson, J. & Krashen, S. (1978). On the domain of the conscious grammar: morpheme orders for corrected and uncorrected ESL student transcriptions. *TESOL Quarterly,* 12, 335-339.

Hu, M. & Nation, P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.

Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Bejoint (eds.): *Vocabulary in Applied Linguistics. London*: MacMillan.

Hulstijn, J. (1996). Does incidental vocabulary acquisition really exist? Paper presented at the11th World Congress of Applied Linguistics (AILA), Jyva¨skyla¨, Finland.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (eds.): The handbook of second language acquisition. Oxford, England: Blackwell, pp. 349-381.

Hulstijn, J. (2005), Theoretical and empirical issues in the study of implicit and explicit second-language learning. Introduction. *Studies in Second Language Acquisition*, 27, 129-140.

Hulstijn, J. & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, 11, 97-112.

Indrarathne, B. & Kormos, J. (2016). Attentional processing of input in explicit and implicit conditions. *Studies in Second Language Acquisition*, 39(3), 401-430.

Irujo, S. (1986). A piece of cake: learning and teaching idioms. ELT Journal, 40(3), 236-241.

Issa, B., Morgan-Short, K. Villegas, B. & Raney, G. (2015). An Eye-tracking Study on the Role of Attention and its Relationship with Motivation. EUROSLA Yearbook, 15(1), 114-142.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24, 541–577.

Jahan, A. & Kormos, J. (2015). The impact of textual enhancement on EFL learners' grammatical awareness of future plans and intentions. International Journal of Applied Linguistics. 25(1), 47-66.

Jiang, N. (2012). Conducting reaction time research in second language studies. New York: Routledge.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal*, 91, 433–445.

Jones, S. & Sinclair, J. (1974). English lexical collocations. A study in computational linguistics. *Cahiers de lexicologie*, 24, 15-61.

Jullfs, A. & Harrington, M. (2011). Aspects of working memory in L2 learning. Language Teaching, 44(2), 137-166.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.

Keyfetz, J. (1978). Natural and Monitored Sequences by Adult Learners of English as a Second Language. Ph.D. Dissertation. Florida State University.

Keating, G., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, 37, 1–32.

Kecskes, I (2006). On my mind: Thoughts about salience, context and figurative language from a second language perspective. *Second Language Research*, 22, 219-237.

Kim, S. H. & Kim, J. H. (2012). Frequency effects in L2 multi-word unit processing: Evidence from self-paced reading. *TESOL Quarterly,* 46(4), 831–841.

Konopka, A. & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology,* 58(1), 68–101.

Kormos, J., & Sa´fa´r, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. Bilingualism: *Language and Cognition*, 11, 261–271.

Krashen, S. (1977). Some issues relating to the Monitor Model. In H. D. Brown, C. Yorio and R. Crymes (eds.): *On TESOL '77: Teaching and Learning English as a Second Language: Trends in Research and Practice*. Washington: TESOL, pp. 144-158.

Krashen, S. (1982). *Principles and practice in second language acquisition*. London: Pergamon.

Krashen, S. & Pon, P. (1975). An error analysis of an advance ESL learner. *Working Papers on Bilingualism*, 7, 125-129.

Krashen, S., Butler, J., Birnbaum, R., & Robertson, J. (1978). Two studies in language acquisition and language learning. *Review of Applied Linguistics*, 39-40, 73-92.

LaBrozzi, R. (2016). The effects of textual enhancement type on L2 form recognition and reading comprehension in Spanish. Language Teaching Research, 20(1), 75-91.

Lackman, R., Lackman, J. L., & Butterfield, E. C. (1979). Cognitive Psychology and Information Processing:An Introduction. Hillsdale, NJ: Lawrence Erlbaun.

Laufer, B. (1992). How Much Lexis is Necessary for Reading Comprehension? In H. Bejoint and P. Arnaud (eds): *Vocabulary and Applied Linguistics*. London: Macmillan, pp. 126-132.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (eds.): *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge, UK: Cambridge University Press, pp. 20–34.

Laufer, B. & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Applied Linguistics,* 29(4), 694–716.

Laufer, B. & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning,* 54(3), 399 – 436.

Laufer, B., & McLean, S. (2016). Loanwords and vocabulary size test scores: A case of different estimates for different L1 learners. *Language Assessment Quarterly*, 13(3), 202-217.

Lee, J. & Benati, A. (2007). Delivering Processing Instruction in Classrooms and Virtual Contexts: Research and practice. London: Equinox.

Lee, S. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning,* 57(1), 87-118.

Lee, S. & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. Language Teaching Research, 21(1), 118-135.

Leow, R. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) fine-grained analysis. *Applied Psycholinguistics*, 19, 133–159.

Leow, R. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, 22(4), 557-584.

Leow, R. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, 84, 496–509.

Leow, R., Grey, S., Matijuan, S. & Moorman, C. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. Applied Language Learning, 13, 93-108.

Leow, R. & Martin, A. (2018). Enhancing the Input to Promote Salience of the L2. In Salience in second language acquisition. New York, NY: Routledge, pp. 167-186.

Leung, J. & Williams, J. (2011). The implicit learning of mappings between forms and contextually-derived meanings. *Studies in Second Language Acquisition*, 33(1), 33–55.

Leung, J. & Williams, J. (2012). Constraints in implicit learning of grammatical form-meaning connections. Language Learning, 62 (2), 634-662.

Lin, P. (2012). Sound evidence: the missing piece of the jigsaw in formulaic language research. *Applied Linguistics*, 33(3), 342-347.

Link, J. & Weiss, D. (2015). Can Working Memory and Inhibitory Control Predict Second Language Learning in the Classroom? SAGE Open, 5(4), 1-11.

Loewen, S. & Inceoglu, S. (2016). The Effectiveness of Visual Input Enhancement on the Noticing and L2 Development of the Spanish Past Tense. *Studies in Second Language Learning and Teaching*, 6(1), 89-110.

Loewen, S. & Reinder, H. (2011). Key concepts in second language acquisition. Hampshire, UK: Palgrave MacMillan.

Long, M. (1983). Linguistic and Conversational Adjustments to Non-Native Speakers. *Studies in Second Language Acquisition*, 5(2), 177-193.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R.B. Ginsberg & E. Kramsch (eds.): *Foreign Language Research in Crosscultural Perspective.* Amsterdam: John Benjamins, pp.39-52.

Long, M. (2015), Second language acquisition and task-based language teaching. Oxford: Wiley-Blackwell.

Long, M. (2017). Geopolitics, methodological issues, and some major research questions. *ISLA*, 1(1), 7-44.

Long, M. (2020). Optimal input for language learning: genuine, simplified, elaborated, or modified elaborated? *Language Teaching,* 53(2), 169-182.

Macis, M., & Schmitt, N. (2017). Not just small potatoes: Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, 21, 321–340.

Malone, J. (2018). Incidental vocabulary learning in SLA. Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 1-25.

Marsden, E., Thompson, S. & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, 1-44.

Martinez, R. & Murphy, V. (2011). Effect of Frequency and Idiomaticity on Second Language Reading Comprehension. *Tesol Quarterly,* 45(2), 267-290.

Mclaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An Information-processing perspective. *Language Learning*, 33, 135-57.

Mohamed, A. (2017). Exposure frequency in L2 reading. An eye-movement perspective of incidental vocabulary learning. Studies in Second Language Acquisition, 40(2), 269-293.

Moreno, E., Federmeier, K. & Kutas, M. (2002). Switching Languages, Switching Palabras (Words): An Electrophysiological Study of Code Switching. *Brain and Language*, 80(2), 188-207.

Murphy, J., Miller, R. & Hamrick, P. (2018). Impact of prior L2 vocabulary knowledge and declarative memory on incidental vocabulary learning through reading. Paper presented at Second Language Research Forum (SLRF) 2018, Universite du Quebec a Montreal, Quebec, Canada.

Myles, F. & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA. Focusing on psycholinguistic FSs and their identification. *Studies in Second Language Acquisition,* 39, 3-28.

Nation, P. (1990). Teaching and learning vocabulary. Boston: Heinle & Heinle.

Nation, P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review,* 63, 59-82.

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language,* 26(2): 1–16.

Nation, P., & Chung, T. (2009). Teaching and testing vocabulary. In Long, M.H., & Doughty, C, J (eds.): *The handbook of language teaching*. Sussex: Blackwell Publishing, pp. 536, 559.

Nation, P., & Webb, S. (2011). Researching and analyzing vocabulary. Boston: Heinle, Cengage Learning.

Negari, G., Azizi, A. Arani, D. (2018). Investigating the effectiveness of audio input enhancement on ELF learners' retention of intensifiers. *International Journal of Instruction*, 11(1), 123-138.

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21, 298–320.

Norris, J. & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528.

Oppenheim, N. (2000). The importance of recurrent sequences for non-native speaker fluency and cognition. In H. Riggenbach (ed.): *Perspectives on fluency*. Ann Arbor: University of Michigan Press, pp. 220-240.

Overstreet, M. (1998). Text enhancement and content familiarity: The focus of learner attention. *Spanish Applied Linguistics*, 2, 229–258.

Paradis, M., (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (ed.): *Implicit and explicit language learning*. London, Academic Press, pp. 393-419.

Paradis, M. (2004). *Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Pawley, A. & F. H. Syder (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards (ed.): *Language and Communication*. London: Longman, pp. 191-227.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. An eye-tracking study. Studies in Second Language Acquisition, 38, 97-130.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381–402.

Pellicer-Sánchez, A. & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In In A. Siyanova-Chanturia & A. Pellicer-Sanchez (eds.): *Understanding formulaic language: A second language perspective*. New York, NY: Routledge, pp. 153-173.

Pellicer-Sánchez, A. & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22, 31–55

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Llanes, A. & Serrano, R. (2018). L2 reading and reading-while-listening in multimodal learning conditions: an eye-tracking study. ELT Research Papers 18.01.

Perfetti, C., Yang, C., & Schmalhofer, F. (2008). Comprehension skills and word-to-text integration processes. *Applied Cognitive Psychology*. 22, 303–318.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology : General*, 119, 264-275.

Perruchet, P., & Pacteau, C. (1991). Implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology: General*, 120, 112-116.

Peters, E. (2009). Learning collocations through attention-drawing techniques: a qualitative and quantitative analysis. In A. Barfield & H. Gyllstad (eds.): *Researching Collocations in Another Language: Multiple Interpretations*. Basingstoke, UK: Palgrave Macmillan, pp. 194-207.

Peters, E. (2012) Learning German formulaic sequences: the effect of two attention-drawing techniques. *Language Learning Journal,* 40(1), 65–79.

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. (2004). Eye movements and semantic composition. In M. Carreriras, & C. Clifton (eds.): *On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond.* New York: Psychology Press, pp. 35-50.

Pinker, S. (1999). Words and rules: the ingredients of language. New York: HarperCollins.

Pitts, M., White, H. & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, 5(2), 271-275.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25-42.

Read, J. & Nation, P. (2004), Measurement of formulaic sequences. in N. Schmitt (ed.): *Formulaic sequences: acquisition, processing and use*. Amsterdam: John Benjamins, pp. 23-35.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 317–327.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology*, 118(3), 219-235.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning,* 63(3), 595-626.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2013). Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness. In J. Bergsleithner, S. Frota, & J. K. Yoshioka (eds.): *Noticing and second language acquisition: Studies in honor of Richard Schmidt*. Honolulu, HI: University of Hawai'i Press, National Foreign Language Resource Center, pp. 249-269.

Rebuschat, P., Hamrick, P., Sachs, R., Ziegler, N., & Riestenberg, K. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition* ,*37*, 299-334.

Rebuschat, P. & Williams, J. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33, 829-856.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. & Prokasy, W. F (eds.): *Classical conditioning, Vol. 2: Current research and theory*. New York: Appleton-Century-Crofts, pp. 64-99.

Reuterskiöld, C. & Van Lancker Sidtis, D. (2013). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy*, 29(2), 219-231.

Richardson, J. (2007). Measures of short-term memory: a historical review. *Cortex*, 43(5), 635-650.

Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning,* 45, 283-331.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition,* 19(2), 223-47.

Robinson, P., Mackey, A., Gass, S. & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (eds.): *The Routledge handbook of second language acquisition*. New York, NY: Routledge, pp. 247-267.

Rosenthal, R. (1991). Meta- analytic procedures for social research (2nd ed.). Newbury Park, CA: Sage.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, 21, 589–619.

Saragi, T., Nation, P. & Meister, G. (1978). Vocabulary learning and reading. *System*, 6, 72-80.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics,* 11(2), 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, 13, 206-226.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11-26.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on attention and awareness in learning. In R. Schmidt (ed.): *Attention and awareness in foreign language learning*. Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center, pp. 1-63.

Schmidt, R. (2001). Attention. In Robinson, P. (ed.): *Cognition and second language instruction*. Cambridge: Cambridge University Press, pp. 3-32.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In Chan, W. M., Chi, S., Cin, K. N., Istanto, J., Nagami, M., Sew, J. W., Suthiwan, T., & Walker, I. (eds.): *Proceedings of classic 2010, Singapore, December 2-4*. Singapore: National University of Singapore, Centre for Language Studies, pp. 721-737.

Schmidt, R., & Frota, S.N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.): *Talking to learn: Conversation in second language acquisition*. Rowley, MA: Newbury House, pp. 237-326.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.

Schumann, J. (1978). The acculturation model for second-language acquisition. In R. Gingras (ed.): *Second-Language Acquisition and Foreign Language Teaching*. Arlington, Virginia: Center for Applied Linguistics, pp. 27-50.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behaviour. *Studies in Second Language Acquisition*, 15, 147–63.

Sharwood-Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics,* 2, 159-168.

Sharwood-Smith, M. (1991). Speaking to many minds: on the relevance of different types of language information for the L2 learner. *Second Language Research*, 7(2), 118-132.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition,* 15, 165-179.

Simard, D. & Foucambert, D (2013). Observing noticing while reading in L2. In J.M. Bergleithner, S.N. Frota & J.K. Yoshioka (eds.): *Noticing and second language acquisition: Studies in honor of Richard Schmidt*. Honolulu: National Foreign Language Resource Center, University of Hawaii at Manoa, pp. 207-226.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11(2), 285-301.

Siyanova-Chanturia, A. & Martinez, R. (2014). The idiom principle revisited. *Applied Linguistics*, 1-22.

Siyanova-Chanturia, A. & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (eds.): *Understanding formulaic language: A second language perspective*. New York, NY: Routledge, pp. 38-61.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27, 251–272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 37(3), 776-784.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E. & van Heuven, W. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, 175, 111-122.

Sonbul, S. & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. Language Learning, 63(1), 121-159.

Spina, S. (2010). The Dictionary of Italian Collocations: Design and Integration in an Online Learning Environment. Conference: Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta.

Spinelli B. & Parizzi F. (2010). Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2. Firenze: La Nuova Italia/ RCS Libri.

Sprenger, S., Levelt, W. & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language,* 54, 161–184.

Stafford, C. & Covitt, G. (1978). Monitor use in adult second language production. *Review of Applied Linguistics*, 39-40, 103-125.

Stubbe R., Nakashima K. (2019). Examining the Effect of Test Order on Passive Recall Versus Active Recall Vocabulary Tests. 2018 BAAL Conference Proceedings, pp. 98-100.

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics,* 30, 1229-1261.

Szudarski, P. (2012). Effects of meaning- and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching and Research,* 1(2), 3–37.

Szudarski, P., & Carter, R. (2014). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, 26(2), 245-265.

Tabossi, P., Fanari, R. & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, 37(4), 529-40.

Tekmen, E. & Daloglu, A. (2006). An Investigation of Incidental Vocabulary Acquisition in Relation to Learner Proficiency Level and Word Frequency. Foreign Language Annals, 39(2), 220-243.

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: a multi-dimensional approach. *Innovation in Language Learning and Teaching*, 12(3), 274-288.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition*, 16, 183-204.

Toomer, M. & Elgort, I. (2019). The Development of Implicit and Explicit Knowledge of Collocations: A Conceptual Replication and Extension of Sonbul and Schmitt (2013). *Language Learning*, 69 (2), 405-439.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613.

Truscott, J. & Sharwood Smith, M. (2011). Input, Intake, and consciousness. The Quest for a Theoretical Foundation. *Studies in Second Language Acquisition*, 33, 497-528.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning,* 69(3), 559-599.

Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it. An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.): *Formulaic sequences: acquisition, processing and use.* Amsterdam: John Benjamins, pp. 153-172.

VanPatten, B. (2004) (ed.). Processing Instruction. NJ Erlbaum.

VanPatten, B. (2006). Input processing. In B. Van Patten & J. Williams (eds.): *Theories in Second Language Acquisition: An Introduction*. Mahwah, NJ: Lawrence Erlbaum, pp. 115–136.

VanPatten, B. (2015). Foundations of processing instruction. International Review of Applied Linguistics in Language Teaching, 53(2), 135-165.

VanPatten, B. (2016). Why explicit knowledge cannot become implicit knowledge. *Foreign Language Annals*, 30, 1-9.

VanPatten, B. & Benati, A. (2010). Key terms in second language acquisition. New York, NY: Continuum.

VanPatten, B., & Cadierno, C. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225-241.

VanPatten, B., & Rothman, J. (2014). Against "rules." In C. Laval & M. J. Arche (eds.): *The grammar dimension in instructed SLA: Theory, research, and practice*. London: Continuum Press, pp. 15-36.

VanPatten, B., & Rothman, J. (2015). What does current generative theory suggest about the explicit-implicit debate? In P. Rebuschat (ed.): *Explicit and implicit learning of languages*. Amsterdam: John Benjamins, pp.91-116.

VanPatten, B., Keating, G., & Leeser, M. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism*, 2, 109–140.

VanPatten, B., Smith, M. & Benati, A. (2020). *Key questions in second language acquisition*. Cambridge: Cambridge University Press.

Vilkaite, L. (2017). Incidental acquisition of collocations in L2 Effects of adjacency and prior vocabulary knowledge. International Journal of Applied Linguistics, 168(2), 248-277.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15, 130–163.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65.

Webb, S. & Chang, A. (2020). How does mode of input affect the incidental learning of collocations? Studies in Second Language Acquisition, 1-22.

Webb, S. & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? Language Teaching Research, 19(6), 667-686.

Webb, S. & Chang, C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review*, 68, 276–290.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43, 55–77.

Webb, S. & Kagimoto, E. (2011). Learning Collocations: Do the Number of Collocates, Position of the Node Word, and Synonymy Affect Learning? *Applied Linguistics*, 32(3), 259–276.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91–120.

Whong, M., Gil, H-G., & Marsden, E. (2014). Beyond paradigm: The 'what' and the 'how' of classroom research. *Second Language Research* 30(4), 551-568.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition,* 27(2), 269-304.

Williams, J. N. (2010). Initial incidental acquisition of word order regularities: Is it just sequence learning? *Language Learning*, 60(S2), 221–24.

Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee, 2007, with eye-movement data. *Studies in Second Language Acquisition*, 35, 323–352.

Woods, D., Kishiyama, M., Yound, W., Herron, T., Edwards, B., Poliva, O., Hink, R. & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. Journal of Clinical and Experimental Neuropsychology, 33(1), 101-111.

Wray, A. (2002). Formulaic language and the lexicon. Cambridge: Cambridge University Press.

Wray, A. (2008). Formulaic language: pushing the boundaries. Oxford: Oxford University Press.

Wulff, S. (2019). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (eds.): *Understanding formulaic language: A second language perspective*. New York, NY: Routledge, pp. 19-37.

Yi, W., Lu, S. & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, 1-31.

Zanjan, F. (2017). The Impact of Teachers' Aural Input Enhancement vs. Textual Enhancement in Learners' Awareness of Ungrammatical Forms. *International Journal of English Linguistics*, 7(5), 190-195.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Addison-Wesley Press.

Appendices

Appendix A. Reading text

La ricetta segreta

秘密的配方，秘方

Da sinistra: lo straniero, Clara, il signor Barbarano, Enrico, la madre di Enrico, Camilla, il signor Ferrarini.

Capitolo 1. In campagna

Mi chiamo Enrico Villa. Mi presento: ho i capelli neri e gli occhi verdi. Sono alto e magro. Mamma e papà dicono che sono una **testa calda**, perché mi arrabbio spesso e se decido di fare qualcosa, nessuno mi può fermare.

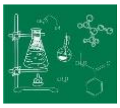
Sono nato e ho abitato in campagna fino all'età di 19 anni. La mia famiglia ha una fattoria in Trentino, una grande casa con molti animali, prati, alberi e campi tutto intorno. Quando ero piccolo giocavo e ero contento: potevo giocare con gli animali, correre nei prati e nei boschi, salire sugli alberi ed ero così felice da **toccare il cielo**.

Alla fattoria c'erano solo mia madre, mio padre e Franco, il loro **braccio destro**. Franco aiutava i miei genitori in tutto: dava il cibo agli animali, stava attento ai campi e agli alberi, e qualche volta giocava con me.

Quando sono diventato più grande però ho iniziato ad annoiarmi: non mi divertivo più. Sempre da solo con i miei genitori e Franco, soprattutto d'estate, quando non c'era la scuola. Cosa potevo fare da solo in campagna? Giocare, fare passeggiate, correre e a volte andare in bicicletta.

Una volta, proprio con la bicicletta, sono andato vicino alla casa del signor Barbarano. La sua casa è la più vicina alla mia, è circa 1 km di strada dalla mia casa.

Il signor Barbarano non ha campi, non ha animali, ma ha una grande serra (温室) dove coltiva i frutti di bosco: mirtilli, more, lamponi, fragole. Sono molto buoni e li va a vendere al mercato del paese: sono molto cari, **costano un occhio**! 

Il signor Barbarano è un po' strano. Ha comprato la casa circa otto anni fa. Viene dalla città: lì lavorava come ingegnere chimico. Mio papà dice che la chimica (化学) è **aria fritta**, niente di importante, solo tante parole, tanta matematica, ma non serve a niente. Invece a me la chimica piace tantissimo. Una volta ho cercato su internet il nome del signor Barbarano: wow, escono tantissimi link! Deve essere un chimico importante! 

Allora sono andato a casa sua per conoscerlo e parlare con lui: finalmente una persona che ama la chimica come me! Ma quando sono arrivato a casa sua, ho avuto una brutta sorpresa, proprio una **doccia gelata**: è uscito di casa all'improvviso, con un bastone in mano!

"Cosa ci fai qui?" ha gridato.

"Mi scusi, mi scusi….volevo solo parlare. Sono…"

"So chi sei, il figlio dei vicini. Vai via! Non **mettere il naso** in casa mia!"

"Io, però…"

Lui ha alzato il bastone.

"Ok, ok" ho detto: c'era una **brutta aria**, sembrava pericoloso (危险).

E ho preso la bicicletta per andare via. Ma poi lui mi ha chiesto:

"Di cosa volevi parlare?"

"Di chimica! Anche a me piace molto, faccio esperimenti (实验) "

"Stai scherzando, ragazzino?" Ha detto lui.

"Ma no! È vero! Mi piace tanto la chimica, ma i ragazzi della mia età non capiscono, e secondo i miei genitori non è importante. Allora ho pensato che qui con Lei…"

"Ah!" Ha detto "Ti piace la chimica? Allora devi saper rispondere a queste semplici domande: qual è la formula (化学配方) dell'acqua?"

Quella domanda era veramente facile! Ho risposto subito, non volevo **perdere la faccia**, fare una brutta figura con lui! Allora mi ha fatto altre domande più difficili e ho risposto a tutto. Allora lui ha abbassato il bastone.

"Puoi entrare in casa" ha detto.

Qui abbiamo parlato, mi ha fatto vedere il suo laboratorio, e siamo diventati amici. Abbiamo passato tanti pomeriggi insieme a fare esperimenti e ho capito che è un genio (天才) della chimica!

Il tempo passato con lui mi ha **aperto gli occhi**: ho capito che all'università voglio studiare chimica!

Così mi sono iscritto all'università di Trento.

Ora abito a Trento, ma nel week end e nelle vacanze torno a casa, in campagna. Mi piace stare con la mia famiglia, e anche vedere il signor Barbarano. Ci scriviamo spesso per email. La sua ultima email era così:

"Ti aspetto sabato alle cinque a casa mia. Ho una cosa da farti vedere. Importante. Ciao."

Sabato allora torno a casa dai miei genitori, e aspetto le cinque per andare dal signor Barbarano.

Capitolo 2. Dov'è il signor Barbarano?

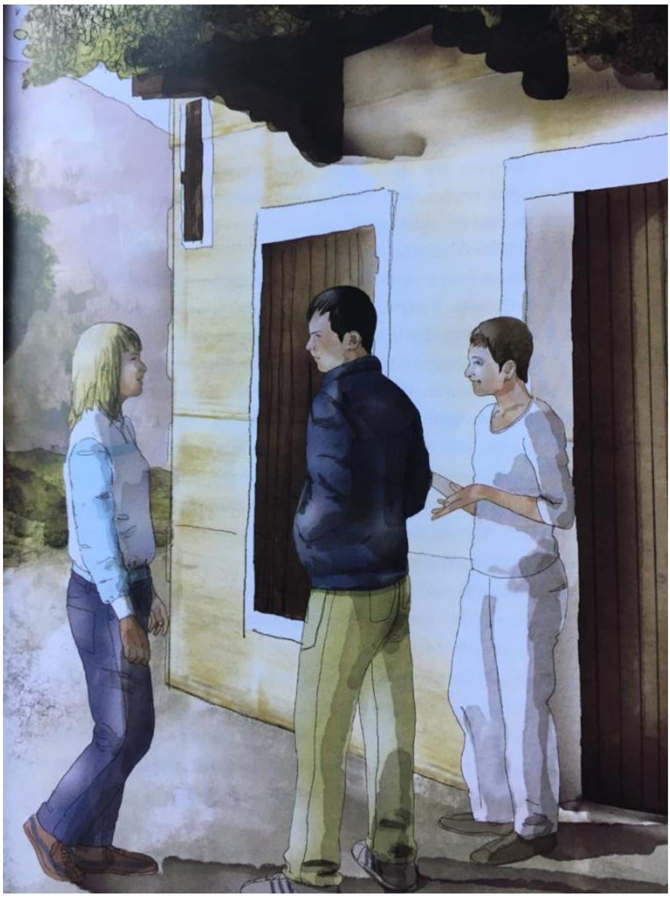
Sono davanti alla casa del signor Barbarano. Suono il campanello della porta. Aspetto. Nessuno viene ad aprire. Suono ancora, ma non viene nessuno. Chiamo il signor Barbarano con il cellulare. È spento, non risponde nessuno! Giro intorno alla casa e lo chiamo:

"Signor Barbarano, signor Barbarano!"

Niente, non risponde. Vado più vicino perché voglio guardare dentro la casa, ma tutte le finestre sono chiuse e non posso.

Dove può essere? È molto strano, lui è sempre in orario. Qui c'è una **brutta aria**, tutto questo non mi piace. Cosa posso fare? Torno a casa.

A casa…oh no! C'è un'altra **doccia gelata**, proprio una brutta sorpresa. Mia madre sta parlando con Camilla. È la figlia di Franco, il **braccio destro** dei miei genitori. Franco aiuta sempre i miei genitori in tutto, e quindi Camilla è spesso a casa nostra. Ha la mia età, da bambini andavamo a scuola insieme. Ma a me non piace. È troppo perfetta: bella, la più brava della classe a scuola. Piace a tutti, ma a me no. È noiosa, e spesso **mette il naso** nella mia camera, entra senza permesso.



"Ciao Enrico!" dice Camilla.

"Ciao Camilla, cosa fai qui?" chiedo.

"I miei genitori sono in vacanza, ma io devo studiare. Quindi sto qui a casa tua per qualche giorno. Tua madre ha detto che devi stare con me!"

"Ma io sono molto occupato" rispondo, e vado in camera mia.

Provo a chiamare ancora il signor Barbarano, ma il telefono è sempre spento. Sono preoccupato (担心): dove può essere?

Dopo poco entra in camera mia madre, molto arrabbiata: "Sei stato antipatico con Camilla! È nostra ospite e non vogliamo fare brutte figure e **perdere la faccia** con i suoi genitori. Devi essere gentile e stare con lei!"

"Mamma, questa è **aria fritta**, non è niente di importante! Io dei grandi problemi e non ho tempo per Camilla!" rispondo, arrabbiato.

"Sei una **testa calda**, come sempre. Ti arrabbi subito. Ma non importa: domani stai con Camilla e basta" Dice mia mamma, e va via.

La mattina dopo è domenica. Mi sveglio molto presto, alle 7. A quest'ora tutti dormono, così posso andare via da casa da solo. Passo in cucina senza fare rumore, sto uscendo quando sento una voce:

"Enrico, dove vai?"

Oh, no, è Camilla!

"Perché sei già sveglia?" le chiedo.

"Mi sveglio sempre presto. E tu, dove stai andando così presto?" Mi chiede.

"Non te lo dico" rispondo, e esco di casa.

"Vengo con te" dice.

"No. Non puoi"

"Sì, posso. Tua madre ha detto che devi stare con me."

" E va bene, vieni. Prendiamo le biciclette" 

Io prendo la mia bicicletta, e do a Camilla la bicicletta di mio padre.

"Stai attenta con questa bicicletta, mio padre l'ha pagata molti soldi, **costa un occhio**" le dico.

"Non ti preoccupare" risponde lei sicura.

Vado molto veloce con la bicicletta, e anche lei. E' molto brava con la bicicletta di mio padre, e non è stanca: deve essere molto sportiva.

Arriviamo a casa del signor Barbarano in pochi minuti. Suono alla porta, richiamo con il cellulare, faccio il giro della casa, ma lui proprio non c'è. Sono davvero preoccupato.

"Chi cerchi?" mi chiede Camilla.

"Un amico che abita qui. Avevo appuntamento con lui ieri, ma non è venuto e non risponde al cellulare"

"Sei preoccupato?" mi chiede.

"Adesso sì. Forse è successo qualcosa di brutto."

“Dovresti chiedere aiuto” mi dice Camilla.

Non mi piace quando Camilla dice cosa devo fare io, ma ha ragione, mi ha **aperto gli occhi**: devo chiamare la polizia.

“Io vado alla polizia, tu vai a casa e stai con i miei genitori: non devono sapere dove sono io.” Dico a Camilla.

“Perché non devono sapere dove sei?” mi chiede

“Perché ai miei genitori non piace il signor Barbarano, non sanno che siamo amici” rispondo.

“Va bene, ci penso io” dice Camilla, e torna a casa con la bicicletta di mio padre.

Finalmente è andata via e sono da solo! Sono così contento che posso **toccare il cielo**!

Vado alla polizia.

“Un mio amico è scomparso, mi dovete aiutare”

“Da quanto tempo è scomparso?” mi chiede il poliziotto

“Da ieri pomeriggio” spiego.

“Noi non possiamo fare niente se è scomparso da meno di 24 ore. Torna domattina” dice il poliziotto.

Oh no!

Suona il mio cellulare: è Camilla.

“Ho detto ai tuoi genitori che sei andato a fare jogging. Ora loro sono andati via, in città. Sono fuori per tutto il giorno”

“Benissimo. Io torno a casa del signor Barbarano. Tu però stai a casa mia!” dico.

“No, vengo con te da Barbarano” risponde lei.

Capitolo 3. A casa del signor Barbarano

Siamo davanti alla casa del signor Barbarano.

“Dobbiamo entrare” dico.

“Sì, ma come?” chiede Camilla.

“Facciamo il giro della casa e vediamo” propongo.

Giriamo intorno alla casa e guardiamo le finestre. Sono tutte chiuse, solo una è aperta, al primo piano sotto il tetto.

“Sì, però è molto in alto. Chi può arrivare fino a lì?” chiedo.

“Io!” esclama Camilla.

Non ho neanche tempo di rispondere: sale sul muro come un gatto, fino alla finestra.

Wow! Questa ragazza è davvero più sportiva di me, perdo la faccia con lei.

“Vieni alla porta sotto, ti apro!” mi dice.

Vado alla porta e lei è già lì. È molto veloce! Forse mi sono sbagliato su di lei, non è così noiosa come pensavo.

Appena entriamo, una vera doccia gelata: tutta la casa è in disordine. I cassetti sono aperti, vestiti e oggetti sono tutti per terra, le sedie e i mobili sono rovesciati per terra.

“Qualcuno è venuto qui e ha messo il naso dappertutto!” esclamo

"Forse qualcuno era arrabbiato con il signor Barbarano? Sai che lui è una testa calda, litiga con tante persone, si arrabbia spesso. Forse qualcuno è venuto qui e hanno litigato" dice Camilla

"Forse, ma mi sembra molto strano!" rispondo.

Andiamo in ogni stanza: dappertutto ci sono disordine e caos. Torniamo nello studio.

"Non c'è il computer. Il suo computer costa un occhio, è molto caro, forse qualcuno è entrato per rubare!" dico.

"Secondo me non volevano rubare il computer" risponde Camilla "Hanno preso anche i quaderni di appunti! Ma che cosa cercavano secondo te?"

"Non lo so, il signor Barbarano mi ha scritto che voleva farmi vedere una cosa. Forse ha inventato o scoperto (发现)qualcosa." rispondo.

"E' un inventore(发明者)?" chiede Camilla.

"Sì, nella chimica." rispondo.

"Ma allora è chiaro! Forse ha inventato qualcosa di nuovo e importante, e qualcuno lo ha portato via, lo ha rapito (绑架） per avere la formula!" dice Camilla.

"Ma certo, ora ho capito! Mi hai aperto gli occhi! E' sicuramente così." Dico.

Cerchiamo ancora in casa, forse c'è qualcosa che ci aiuta a capire chi ha rapito il signor Barbarano.

Andiamo in cucina: sul tavolo c'è un quaderno di ricette, e Camilla inizia a leggerlo. "Guarda, è scritto a mano, e ci sono ottime torte!" dice Camilla.

"Sei anche una cuoca?! Senti Camilla, non sappiamo dov'è il signor Barbarano, non sappiamo se sta bene, non sappiamo cosa è successo qui: non è il momento di leggere le ricette delle torte, è aria fritta!" mi arrabbio.

"Ma no, guarda, è molto strano! Nelle prime pagine ci sono le ricette, e poi ci sono scritte che non capisco" dice Camilla.

"Ma ti sembra importante?" chiedo, ancora arrabbiato.

"Certo! Ho capito!" esclama Camilla "Le parole sono scritte al contrario. Anche Leonardo da Vinci scriveva così. Da destra verso sinistra. Così nessuno può capire cosa scrivi!

Guarda qui, sembrano formule* di chimica!"

"Camilla…bravissima! Hai ragione, è proprio vero! Forse qui c'è il segreto della scoperta del signor Barbarano!"

Ora Camilla è proprio il mio braccio destro in tutta questa storia, mi sta aiutando tantissimo!

Usciamo dalla casa del signor Barbarano. Vogliamo portare a casa il quaderno di ricette per leggerlo bene.

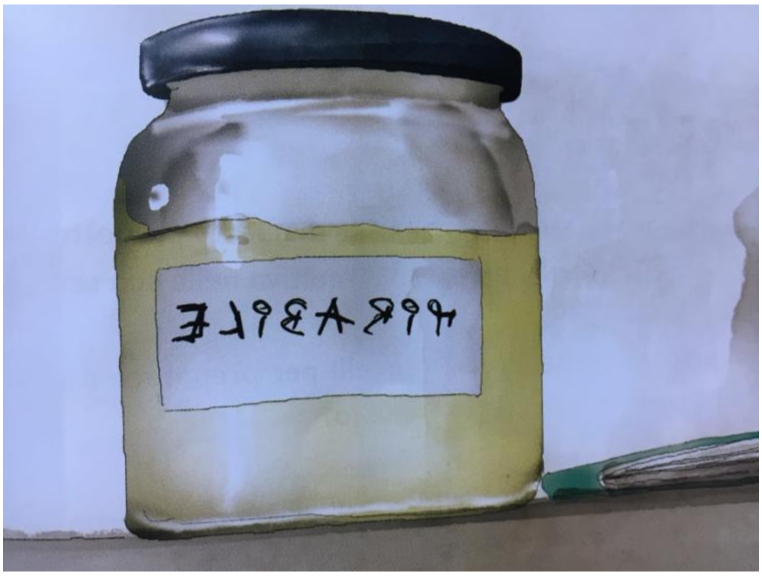
Camilla si gira.

"Cosa c'è? Hai sentito qualcosa?"

"Non lo so, forse, un rumore, sembrava una persona. Andiamo via, qui c'è una brutta aria, ho paura."

Quando arriviamo a casa siamo così contenti che ci sembra di toccare il cielo: siamo al sicuro!

Capitolo 4. Mirabile!



Ora siamo a casa. I miei genitori, per fortuna, non sono ancora tornati.

Io e Camilla andiamo subito in camera mia e ci sediamo alla scrivania. Con attenzione, leggiamo le parole dal quaderno di ricette del signor Barbarano. Le parole sono scritte al contrario, quindi è difficile leggere. Camilla legge a voce alta le parole al contrario, e io le scrivo giuste. Dopo, provo anche io a leggere, ma perdo la faccia: lei è molto più brava e veloce di me!

Parola dopo parola, ri-scriviamo nel verso giusto tutto il quaderno, e poi leggiamo tutto. Ci sono ingredienti, formule e istruzioni su cosa fare. Spesso parla di un "prodotto" (美容产品)

"Parla molto spesso di frutti di bosco, come mirtilli, lamponi, fragole: forse sono l'ingrediente più importante di questo prodotto" dice Camilla.

"Il Signor Barbarano ha molte piante di frutti di bosco! Lui cura le piante e poi vende i frutti al mercato: **costano un occhio** perché sono i più buoni della città!" rispondo.

"Quindi il signor Barbarano ha usato i frutti di bosco per questo prodotto. Qui dice che si chiama 'Mirabile'" dice Camilla.

"Qui dice anche che devi mettere questo prodotto sulla pelle del viso, della faccia una volta al giorno.

Forse è una crema per la pelle. " continuo io. 

"Non lo so. Secondo me è non è solo una crema di bellezza. Deve essere qualcosa di più. Qualcosa di nuovo, straordinario, speciale." Dice Camilla.

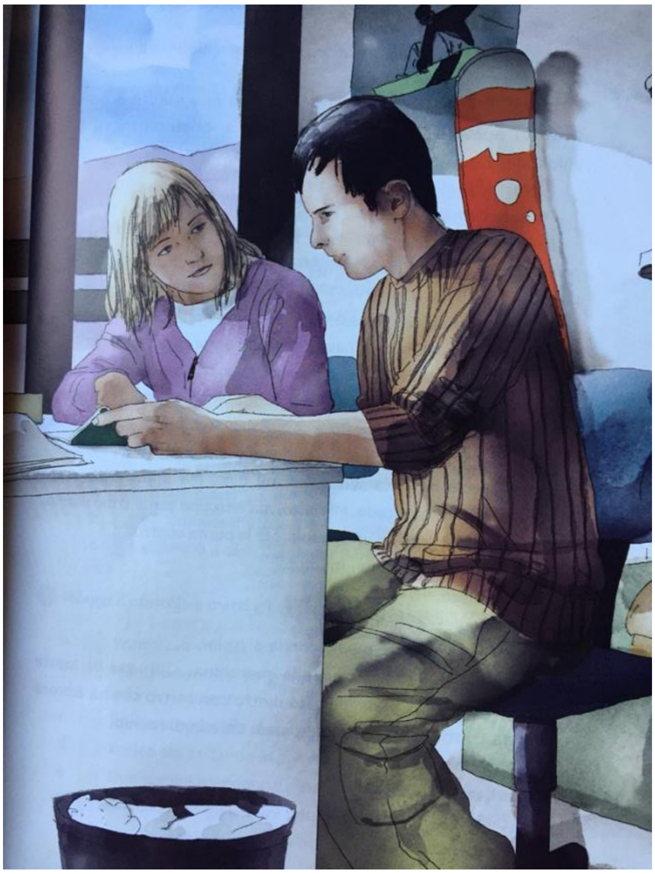
"Ma certo, hai ragione!" dico io "Secondo me è una crema molto speciale: quando la metti, la pelle sembra molto più giovane."

Ora ho aperto gli occhi, ho capito: il signor Barbarano ha creato questa nuova crema speciale, e qualcuno vuole rubare (偷) la formula!

"Ci sono altre pagine, continuiamo a leggere?" chiede Camilla.

"No, non serve, ho capito cosa succede. Sai che sei proprio brava Camilla?" rispondo.

"Brava?"



"Sì, hai capito che nel quaderno non c'erano solo ricette, leggi al contrario molto velocemente, sei davvero intelligente come dice mia madre!"

"Sì, ma ai ragazzi questo non piace. Neanche a te io piaccio, giusto?" risponde Camilla.

"E' vero, prima non mi piacevi. Perché sei troppo perfetta. Ma ora ho capito che erano tutte cose stupide, aria fritta. Ora mi piaci! Ma non c'è tempo per questo, voglio andare alla polizia a dire tutto quello che abbiamo capito." Dico io.

"Allora andiamo subito!" dice Camilla.

"No, Camilla, tu stai qui per favore! Se arrivano i miei genitori e a casa non c'è nessuno, loro cominciano a telefonare e fare domande. È meglio se quando loro tornano, tu sei a casa."

"Va bene, allora rimango qui." Risponde Camilla.

"Grazie Camilla, sei davvero il mio braccio destro! Il tuo aiuto è molto importante!"

Esco di casa. Penso al signor Barbarano. Non so se sta bene, non so dov'è, ho paura (害怕) per lui!

Cammino veloce sulla strada per il paese. Come al solito, non ci sono macchine. Vicino alla strada c'è un bosco. 

Poi, una doccia gelata. Una macchina arriva molto veloce, si ferma davanti a me. Due uomini escono dalla macchina. Resto fermo in piedi per qualche secondo. Qui c'è davvero una brutta aria, hanno anche le pistole  ! Comincio a correre verso il bosco. Loro corrono dietro di me per prendermi. Continuo a correre. Sono alto e veloce, più veloce di loro, non possono prendermi!

Penso di essere lontano da loro e sono così felice che potrei toccare il cielo, invece mi sbaglio! Non sono lontani, sono proprio dietro di me! Così, quando mi fermo arrivano e mi prendono.

"Adesso vieni con noi. Abbiamo visto che hai messo il naso in casa del signor Barbarano."

"Cosa volete?" chiedo.

"Stai zitto." Dice quello con la pistola.

Di solito sono una testa calda: nessuno può dirmi di stare zitto e faccio solo quello che voglio io. Ma adesso ho molta paura di quella pistola! Così, sto zitto e vado con loro fino alla macchina. Un uomo guida, e quello con la pistola si siede dietro vicino a me.

"Dove andiamo?" chiedo.

"Stai zitto" ripete quello con la pistola.

Ho davvero molta paura!

Capitolo 5. Rapito

Il viaggio in macchina dura poco, solo 15 minuti. La macchina si ferma davanti a una vecchia e grande casa, nel bosco. Scendo dalla macchina con i due uomini e entriamo nella casa.

L'uomo con la pistola mi porta in una stanza buia, poi esce e chiude a chiave. Rimango al buio, provo a guardare intorno ma non vedo niente. Poi però sento una voce:

"Enrico, sei proprio tu?"

E' il signor Barbarano! Sono contento di averlo trovato! E sta bene! Anche se siamo chiusi in questa stanza buia, posso toccare il cielo!

"Enrico, cosa fai qui?" mi chiede il signor Barbarano.

Gli racconto tutto. Lui mi abbraccia, come un padre.

"Mi dispiace tanto. Adesso anche tu sei qui, con questi criminali (犯人). C'è una brutta aria, ho paura per te"

Dopo un po' di tempo, i miei occhi iniziano a vedere meglio al buio. Così posso vedere la stanza e anche il signor Barbarano. Ma c'è qualcosa di strano!

"Signor Barbarano, cosa ha fatto alla sua faccia? È strana!"

"Cosa vuoi dire? Perché strana?"

"E' più…giovane!"

"Si può vedere anche così al buio!"

"E' la sua crema Mirabile, vero?" chiedo.

"Mirabile! Come sai questo nome?"

"Io e la mia amica Camilla abbiamo trovato il suo quaderno, e abbiamo capito cosa ha scritto, signor Barbarano."

"Siete stati più intelligenti di loro. Hanno messo il naso in tutta la mia casa, hanno guardato e hanno cercato, ma non hanno trovato il quaderno." Dice il signor Barbarano. "Ma chi sono queste persone?"

"Sono persone cattive, Enrico, criminali"

"Ma cosa è successo? Mi racconti tutto!"

Il signor Barbarano racconta:

"Sto studiando i frutti di bosco da 3 mesi. Dopo molto lavoro, ho capito come fare una nuova crema: Mirabile. Ho usato la crema ogni giorno sulla faccia per una settimana, e guarda: sembro molto più giovane! Un risultato incredibile!

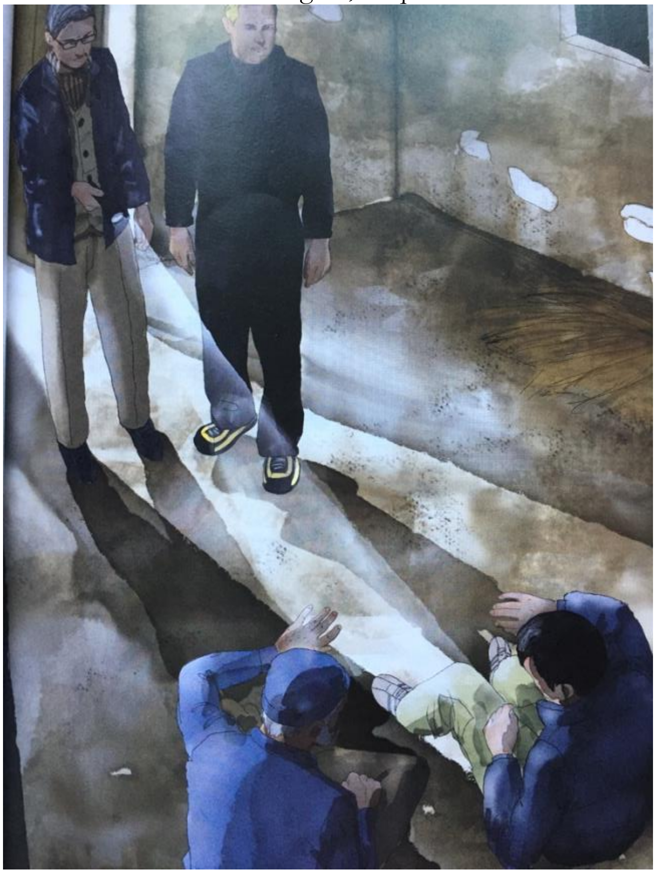
Allora ho pensato di raccontare a qualcuno di questa nuova crema. Ho chiamato il signor Ferrarini. È un mio vecchio collega, una brava persona e anche un mio amico, pensavo. Invece era tutta aria fritta: è una persona cattiva, un criminale.

Quando gli ho scritto, mi ha riposto: "Grazie, ti chiamo presto"

Ma qualche giorno dopo ho avuto una terribile doccia gelata. Non ha chiamato: è venuto a casa mia, voleva la formula per fare la crema. Non gli ho dato la formula, e così mi ha portato qui. Dicono che devo dargli la formula, oppure mi fanno del male."

"Oh no! Quindi Ferrarini ha fatto tutto questo! Ma era suo amico?"

"Credevo di sì, ma ora ho aperto gli occhi, è solo un criminale! Ha capito che la crema Mirabile può costare un occhio nei negozi, e quindi vuole la formula per fare la crema. Così la può vendere e guadagnare molti soldi!".



In quel momento si apre la porta e entra un uomo. È l'uomo che mi ha portato qui.

"Adesso porto via il ragazzo, Barbarano" dice uno.

"No, Ferrarini!" grida lui.

"Non hai voluto dirci niente, adesso sentiamo cosa può dire il ragazzo. Lo porto dal mio braccio destro Franz. Mi aiuta sempre con questi problemi. È una testa calda, si arrabbia subito se non rispondi alle sue domande, ed è molto forte…"

Molto forte?? Si arrabbia?? Cosa vuole farmi?? Non dico niente per non perdere la faccia, ma ho molta paura.

"Lasciate qui il ragazzo! Lui non sa niente!" dice il signor Barbarano.

"Allora parla, Barbarano! Ci devi dire dov'è la formula per la crema!"

Barbarano non parla.

"Barbarano! Porto il ragazzo da Franz!"

"No, no, va bene, parlo. La formula è a casa mia."

Capitolo 6. Dov'è Camilla?

Il signor Ferrarini guarda l'altro uomo, il suo braccio destro. Lui parla per la prima volta. Non sembra italiano, forse è tedesco.

"Non è vero" dice "è tutta aria fritta. Barbarano vuole che perdiamo tempo. Abbiamo guardato in tutta la casa. In ogni cassetto, in ogni angolo. La formula non è a casa."

"Invece è a casa mia! È scritta in un quaderno" Dice Barbarano.

"E dov'è questo quaderno?" chiede Ferrarini

"In cucina, sul tavolo."

"Sul tavolo?! Non è possibile!" il tedesco è molto arrabbiato. Sa che ha perso la faccia, perché non ha capito che la formula era proprio davanti a lui, sul tavolo della cucina.

"Adesso non perdiamo tempo! Torniamo a casa di Barbarano e prendiamo il quaderno" dice Ferrarini.

Capisco che devo dire qualcosa. Se questi due uomini vanno a casa del signor Barbarano e non trovano il quaderno, si arrabbiano davvero. E non sappiamo cosa possono fare se sono davvero arrabbiati. Allora dico:

"Veramente, il quaderno non è più a casa del signor Barbarano."

Tutti mi guardano.

"E dov'è?" chiede il tedesco.

"E' a casa mia. Stamattina io e la mia amica siamo andati a casa di Barbarano e abbiamo trovato il quaderno." Rispondo.

"Ho visto che tu e la tua amica eravate da Barbarano stamattina. Avete messo il naso in tutta la casa. E…sì è vero! In effetti la ragazza aveva in mano un quaderno quando siete andati via!" dice il tedesco.

"Quindi il quaderno è a casa tua!" dice Ferrarini.

"Esatto"

"Questa è una doccia gelata. Come possiamo prendere il quaderno adesso?" chiede lo straniero.

"Non ti preoccupare. So come fare. Adesso telefono a Clara e le dico di venire qui subito." risponde Ferrarini. "Tu prendi il ragazzo."

Ci portano fuori dalla stanza buia. Ora sono in una grande cucina. Che cosa succede adesso? Chi è Clara? Cosa vuole fare Ferrarini?

Dopo dieci minuti arriva una donna. È bassa e magra, ha circa 40 anni. Ha vestiti eleganti, sembra una persona normale.

Si avvicina a Ferrarini e gli dà un bacio. Forse sono amici, o anche più che amici.

"Allora? Perché mi hai chiamato?" chiede.

"Devi fare una cosa per noi. Abbiamo bisogno di una donna. Una donna elegante e tranquilla come te, che non sembri una testa calda" risponde Ferrarini.

"Cosa devo fare?" chiede ancora lei.

"Devi andare con questo ragazzo a casa sua. È vicino alla casa di Barbarano. Devi entrare in casa con lui e devi dire che sei un'amica di Barbarano. Lui deve prendere un quaderno e poi deve uscire con te. Devi stare sempre con lui." Risponde Ferrarini. Poi guarda me e dice:

"E tu, ragazzo: stai attento! Se non ti comporti bene, se tu dici qualcosa di troppo, il tuo amico Barbarano è morto (死了). Capito??"

"Ho capito, ho capito!" rispondo subito.

Ferrarini vede che ho paura. Allora mi dice:

"Non ti devi preoccupare. Prendi il quaderno e vieni via. Quando abbiamo il quaderno ti lasciamo andare libero."

"E il signor Barbarano?" chiedo io.

"Il signor Barbarano viene con noi. Ci deve aiutare a fare la crema Mirabile. E poi lasciamo libero anche lui. Noi non siamo cattivi. Vogliamo solo guadagnare molti soldi, e la crema Mirabile costerà un occhio nei negozi. Così saremo ricchi!" dice Ferrarini.

Il signor Barbarano si arrabbia molto e dice: "Ferrarini, dici che non siete criminali?! Mi avete portato qui, e adesso avete anche preso questo ragazzo! Devi aprire gli occhi Ferrarini, tu sei davvero un criminale!"

Ferrarini non gli risponde.

Esco con la donna. Saliamo in macchina. Lei non dice una parola per tutto il tempo, è molto seria. C'è proprio una brutta aria, io ho paura per me e per il signor Barbarano.

Arriviamo a casa mia. La donna mi dice:

"Hai capito, vero? Non devi dire una parola. Se non fai come dico io, chiamo il mio amico Ferrarini e Barbarano è morto."

"Ho capito" rispondo.

Appena entro, arriva mia madre:

"Ma dove sei stato?" comincia a parlare. Ma poi si ferma perché vede la donna.

"Buongiorno signora" dice la donna.

"Buongiorno" risponde mia madre.

"Mi chiamo Cristina Vadegotti e lavoro con il dottor Barbarano. Enrico ha un quaderno del dottor Barbarano, e dobbiamo prenderlo." Dice la donna.

"Capisco, va bene" risponde mia madre.

"Dov'è il quaderno, Enrico? Andiamo a prenderlo" mi chiede la donna.

"Va bene" rispondo.

Andiamo nella mia camera. Il quaderno è sulla scrivania. Lo prendo.

Però Camilla non c'è! Dove può essere? Forse ha capito tutto, ed è andata alla polizia! Questo pensiero mi fa toccare il cielo!

Capitolo 7. Finalmente a casa!

Siamo alla porta di casa. Mia madre ci saluta e mi chiede quando torno. La donna risponde per me:

"Tra due ore, signora."

Attraversiamo il giardino per andare alla macchina. Vedo Camilla, o almeno mi sembra di vedere Camilla. È nascosta (隐藏) dietro un albero, e guarda me e la donna.

Torniamo alla vecchia casa nel bosco dove hanno portato Barbarano.

"Adesso posso andare via?" chiedo.

"Non subito. Fra poco. Ma non aver paura." Risponde la donna. Ma c'è una brutta aria, e io ho molta paura.

Dentro la vecchia casa, ci sono Ferrarini e il tedesco. La donna dà il quaderno a Ferrarini. Lui lo apre e legge. Dopo poco si arrabbia molto e dice:

"Ma che cos'è questo?! Qualche ricetta di torte e dopo strane parole che non si possono capire! È tutta aria fritta!"

Sembra molto arrabbiato, una vera testa calda. Mi fa paura con la pistola in mano, così arrrabbiato. Allora dico subito:

"E' facile da capire! Le lettere e le parole sono scritte da destra a sinistra. Come Leonardo da Vinci."
"Cioè è tutto scritto al contrario?" chiede Ferrarini.
"Sì, esatto." Rispondo.

"Barbarano è intelligente. Ha capito che la sua crema Mirabile può costare un occhio nei negozi. E quindi, ha scritto tutto al contrario così nessuno poteva mettere il naso nella sua formula."

"Questa è una doccia gelata!" dice il tedesco arrabbiato. "Dobbiamo leggere tutte le parole al contrario. Ci vuole troppo tempo!"

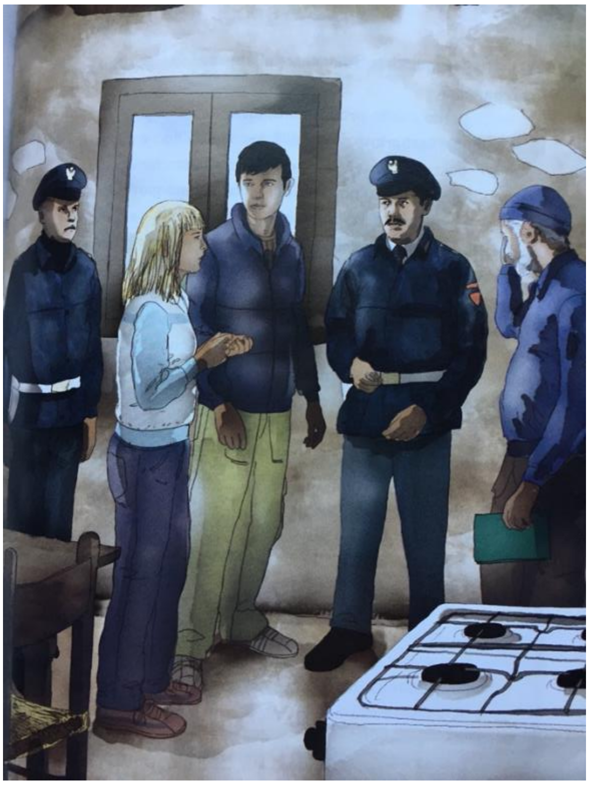
"Non ti preoccupare." Risponde Ferrarini. "Barbarano è intelligente, ma noi abbiamo le pistole. Vai a chiamarlo"

Il tedesco ride e va a chiamare Barbarano: "Ora devi leggere il quaderno per noi."

"Va bene" dice il signor Barbarano.

Il signor Barbarano inizia a leggere, mentre Ferrarini scrive tutto. Dopo poco sentiamo dei rumori.

"Che cosa succede?" chiede il tedesco, e corre fuori dalla casa con la pistola in mano.



Anche Ferrarini prende la pistola, e dice a me e Barbarano:

"State zitti e non fate rumore. Oppure…" ma non può finire la frase, perché nella stanza entrano degli uomini, anche loro hanno la pistola! Sono poliziotti!

Mi sembra di toccare il cielo, siamo salvi!!

Ferrarini lascia cadere a terra la pistola e alza le mani: la polizia porta via lui, il suo braccio destro (il tedesco) e la donna.

Uno dei poliziotti viene da noi: "state bene?"

"Sì, stiamo bene!" rispondo. "Come ci avete trovato?"

"La tua amica Camilla ci ha chiamato e ci ha aperto gli occhi su quello che succedeva qui! Ci ha spiegato tutto e siamo venuti subito" risponde il poliziotto.

In quel momento arriva Camilla.

"Quando ho visto Enrico con quella donna ho capito tutto. Ho preso la bicicletta e ho seguito la macchina. Ho visto Enrico che entrava qui con la donna, e allora ho chiamato la polizia" spiega.

"Grazie, Camilla!" dico, e la abbraccio.

"La tua amica è molto intelligente. Ci ha fatto perdere la faccia: lei ha capito dove era Barbarano prima di noi!" dice il poliziotto.

"Possiamo tornare a casa adesso?" chiedo io.

"Non ancora, dovete venire con noi e raccontare bene cosa è successo". Risponde il poliziotto.

Dopo due ore siamo finalmente a casa. Raccontiamo tutto ai miei genitori, e poi corriamo a casa del signor Barbarano.

Suoniamo alla porta, e questa volta ci apre subito!

"Ragazzi, che bello vedervi! Venite, vi faccio vedere come preparare la crema Mirabile!"

Appendix B. Comprehension test

1. Riordina le immagini (___/6).

2. Rispondi con Vero o Falso (_____/14).

1. Quando Enrico era bambino, non gli piaceva abitare in campagna.	V	F
2. Enrico studia chimica all'università.	V	F
3. Enrico all'inizio è felice di vedere Camilla.	V	F
4. Il signor Barbarano non è a casa.	V	F
5. Camilla entra in casa del signor Barbarano dalla finestra.	V	F
6. Camilla e Enrico trovano un quaderno di ricette in cucina.	V	F
7. Enrico è più bravo di Camilla a leggere il quaderno al contrario.	V	F
8. Enrico incontra due uomini in macchina.	V	F
9. La crema Mirabile fa diventare la pelle più chiara.	V	F
10. Gli uomini con la pistola portano Enrico in una casa in città.	V	F
11. Clara sembra una criminale.	V	F
12. Camilla guarda Enrico mentre lui è con Clara.	V	F
13. Camilla chiama la polizia.	V	F
14. Camilla, Enrico e il signor Barbarano devono parlare con la polizia.	V	F

Appendix C. Self-paced-reading text items and comprehension T/F questions

SET 1 – Target items and comprehension questions

1. Carlo vuole sapere tutto di me, e mette il naso fra le mie cose. Non mi piace!

Carlo mi sta simpatico.

2. L'insegnate ha fatto una domanda facile a Paolo, ma lui non ha saputo la risposta: ha perso il naso davanti a tutti

Paolo ha studiato bene.

3. Ho sempre pensato che Napoli è una bella città. Ma poi sono andata a Napoli e ho aperto gli occhi: è sporca e rumorosa, non mi piace!

Sono andata a Napoli.

4. Volevo comprare un nuovo computer ma non ho potuto: costava un braccio, così uso ancora il vecchio computer

Ho un nuovo computer.

5. Dopo tanto studio, martedì finalmente mi sono laureata: mi sembra di toccare il cielo per la felicità!

Ho finito l'università

6. Alice e Pietro hanno litigato ieri sera alla festa. Dopo c'era una bassa aria: nessuno voleva più parlare.

La festa non è stata divertente.

7. Giorgio si è arrabbiato ieri sera, e ha litigato con due ragazzi davanti al pub. È una testa calda, non va bene.

Giorgio ha passato una serata tranquilla.

8. Questo è Giovanni: mi aiuta nel mio lavoro, è il mio piede destro. Se io non ci sono, puoi chiedere a lui.

Puoi parlare con Giovanni se io non ci sono.

9. Vorrei iscrivermi in accademia ma secondo mio padre studiare arte è come studiare aria fritta. Invece secondo me è importante.

A mio padre non interessa l'arte.

10. Ieri dovevo partire per New York, ma in aeroporto ho avuto un bagno gelato: hanno cancellato il mio volo e non sono partita!

Ieri sono arrivata a New York

SET 2 – Target items (comprehension questions were the same as in Set 1)

1. Carlo vuole sapere tutto di me, e mette i piedi fra le mie cose. Non mi piace!

2. L'insegnate ha fatto una domanda facile a Paolo, ma lui non ha saputo la risposta: ha perso la faccia davanti a tutti.

3. Ho sempre pensato che Napoli è una bella città. Ma poi sono andata a Napoli e ho aperto la testa: è sporca e rumorosa, non mi piace!

4. Volevo comprare un nuovo computer ma non ho potuto: costava un occhio, così uso ancora il vecchio computer.

5. Dopo tanto studio, martedì finalmente mi sono laureata: mi sembra di toccarel'aria per la felicità!

6. Alice e Pietro hanno litigato ieri sera alla festa. Dopo c'era una brutta aria: nessuno voleva più parlare.

7. Giorgio si è arrabbiato ieri sera, e ha litigato con due ragazzi davanti al pub. È una bocca calda, mi fa preoccupare.

8. Questo è Giovanni: mi aiuta nel mio lavoro, è il mio braccio destro. Se io non ci sono, puoi chiedere a lui.

9. Vorrei iscrivermi all'Accademia ma secondo mio padre studiare arte è come studiare aria lessa. Invece secondo me è importante.

10. Ieri dovevo partire per New York, ma in aeroporto ho avuto una doccia gelata: hanno cancellato il mio volo e non sono potuta partire!

Fillers and comprehension questions

11. In inverno, posso mangiare frutti come le fragole perché si usano le serre.

Mi piacciono le fragole

12. Per molti studenti la chimica è una materia difficile, ma a me piace molto.

Per me la chimica è difficile

13. Ho letto sul giornale che il figlio di un ricco industriale è stato rapito.

Stamattina non ho comprato il giornale

14. Gli scienziati devono fare molti esperimenti per trovare nuove medicine.

Inventare nuove medicine è facile

15. Domani ho un esame e non ho ancora studiato tutto: sono molto preoccupato.

Sono pronto per l'esame

16. Leonardo da Vinci è stato un grande inventore: ha creato il primo modello di elicottero.

Leonardo ha studiato il volo

17. Ieri in centro mi hanno rubato la borsa con il portafogli! Mi sono arrabbiata molto, ma per fortuna il cellulare era in tasca!

Ho ancora il mio cellulare

18. Le ragazze spesso hanno paura degli insetti, ma io no!

Se vedo un insetto non sono tranquilla

19. Quando ero bambina e il mio gatto è morto sono stata molto triste. Quindi non voglio più avere gatti.

A casa ho un gatto

20. Volevo mangiare un dolce, ma mio fratello ha nascosto il cioccolato perché vuole mangiarlo tutto lui!

A mio fratello piacciono i dolci

Appendix D. Offline posttests

L2 to L1 translation

Prova 1. Completa la traduzione dal cinese all'italiano. (______/10)

1.

我们谈了一个小时的话，但是什么都决定不了。 我们说的都是废话！

Abbiamo parlato per un'ora, ma non abbiamo deciso niente di importante. Abbiamo parlato solo di A__________________ F__________________!

2.

对我们来说这台电脑太贵了！ 我们买不了，价格高极了！

Questo computer è troppo caro per noi! Non possiamo comprarlo, C__________________ U_______O__________________.

3.

欢迎！ 您好！ 我给您介绍Matteo。 他帮助我做所有的事和解决实际问题:他是我的助理人员。

Buongiorno e benvenuto. Le presento Matteo. Mi aiuta in tutto e si occupa di tutti i problemi pratici: è il mio B__________________ D__________________.

4.

Chiara 跳舞的时候在众人面前跌倒了，她真丢脸了！

Chiara è caduta mentre ballava davanti a tutti. Ha davvero P__________________ L_______ F__________________.

5.

Marco 一直/常常都生气了，他来决定干一个事的时候，谁也不能让他改变主意 他真是一个性急子。

Marco si arrabbia continuamente e quando decide di fare qualcosa non è possibile fargli cambiare idea: è davvero una T__________________ C__________________!

6.

我本来认为Luca是一个好孩子， 但是他对Anna不好。 现在我清楚地明白了: 他不是好孩子!

Prima credevo che Luca fosse un bravo ragazzo, ma è stato molto cattivo con Anna e ora H______ A__________________ G_______ O__________________: non è per niente un bravo ragazzo!

7.

Luigi很嫉妒， 他不要他的妻子发短信给别的男人。所以，为了检查，他每天探听他妻子的手机。

Luigi è molto geloso e non vuole che sua moglie mandi messaggi ad altri uomini. Quindi M__________________ I______ N__________________ nel telefono di sua moglie ogni giorno, per controllare.

8.

我知道了MARCO不再和我们一起工作， 让我感到惊讶。我真的没想到。

Quando ho saputo che Marco non lavora più con noi è stata una brutta sorpresa. Proprio una D__________________ G__________________.

9.

我面试了，得到了工作的位置！我是这么的高兴， 像漫步云端的感觉。

Ho fatto un colloquio e ho avuto il lavoro! Sono così felice che mi sembra di T_________________ I______ C_________________!

10.

昨天晚上，我一个人坐地铁回家了。有坏人让我有点害怕。因此有点危险。

Ieri sera sono tornata a casa da sola in metropolitana. C’erano delle brutte persone e avevo un po’ di paura. Insomma, c’era una B_________________ A_________________.

Form recognition test

Prova 2. Segna la frase più giusta. (_____/10)

1.

a) Dopo un anno molto stressante sono finalmente in vacanza! Sono così felice che mi sembra di prendere il cielo.
b) Dopo un anno molto stressante sono finalmente in vacanza! Sono così felice che mi sembra di avere il cielo.
c) Dopo un anno molto stressante sono finalmente in vacanza! Sono così felice che mi sembra di toccare il cielo.
d) Dopo un anno molto stressante sono finalmente in vacanza! Sono così felice che mi sembra di essere il cielo.

2.

a) Non ho comprato il nuovo smartphone perché costa un orecchio, è davvero troppo caro.
b) Non ho comprato il nuovo smartphone perché costa un braccio, è davvero troppo caro.
c) Non ho comprato il nuovo smartphone perché costa un occhio, è davvero troppo caro.
d) Non ho comprato il nuovo smartphone perché costa una mano, è davvero troppo caro.

3.

a) Chiara dice sempre di essere più brava di Marta nel ping-pong. Ma oggi Marta ha vinto e Chiara ha perso la mano!
b) Chiara dice sempre di essere più brava di Marta nel ping-pong. Ma oggi Marta ha vinto e Chiara ha perso la faccia!
c) Chiara dice sempre di essere più brava di Marta nel ping-pong. Ma oggi Marta ha vinto e Chiara ha perso i capelli!
d) Chiara dice sempre di essere più brava di Marta nel ping-pong. Ma oggi Marta ha vinto e Chiara ha perso la pancia!

4.

a) Marco litiga con tutti, si arrabbia sempre e non pensa molto prima di parlare. È una vera bocca calda.
b) Marco litiga con tutti, si arrabbia sempre e non pensa molto prima di parlare. È una vera mente calda.
c) Marco litiga con tutti, si arrabbia sempre e non pensa molto prima di parlare. È un vero cuore caldo.
d) Marco litiga con tutti, si arrabbia sempre e non pensa molto prima di parlare. È una vera testa calda.

5.

a) Diego aiuta Chiara in tutto, e sono sempre insieme. Diego è il braccio destro di Chiara.
b) Diego aiuta Chiara in tutto, e sono sempre insieme. Diego è la gamba destra di Chiara.
c) Diego aiuta Chiara in tutto, e sono sempre insieme. Diego è la mano destra di Chiara.
d) Diego aiuta Chiara in tutto, e sono sempre insieme. Diego è il piede destro di Chiara.

6.

a) Il mio fidanzato mi ha lasciata oggi. Fino a ieri sembrava felice, è stata una vera pioggia gelata.
b) Il mio fidanzato mi ha lasciata oggi. Fino a ieri sembrava felice, è stata una vera doccia fredda.
c) Il mio fidanzato mi ha lasciata oggi. Fino a ieri sembrava felice, è stata una vera doccia gelata.
d) Il mio fidanzato mi ha lasciata oggi. Fino a ieri sembrava felice, è stata una vera pioggia fredda.

7.

a) Non mi piace uscire con Luca. Non parla mai di cose importanti: parla solo di aria cotta.
b) Non mi piace uscire con Luca. Non parla mai di cose importanti: parla solo di aria bollita.
c) Non mi piace uscire con Luca. Non parla mai di cose importanti: parla solo di aria calda.
d) Non mi piace uscire con Luca. Non parla mai di cose importanti: parla solo di aria fritta.

8.

a) Prima Carlo pensava di essere fidanzato con la ragazza giusta. Ma ora ha aperto la testa: Giulia non va bene per lui, deve cambiare fidanzata.
b) Prima Carlo pensava di essere fidanzato con la ragazza giusta. Ma ora ha aperto gli occhi: Giulia non va bene per lui, deve cambiare fidanzata.
c) Prima Carlo pensava di essere fidanzato con la ragazza giusta. Ma ora ha aperto le orecchie: Giulia non va bene per lui, deve cambiare fidanzata.
d) Prima Carlo pensava di essere fidanzato con la ragazza giusta. Ma ora ha aperto la bocca: Giulia non va bene per lui, deve cambiare fidanzata.

9.

a) Mia madre vuole sempre sapere tutto di me, e quindi mette il naso fra le mie cose.
b) Mia madre vuole sempre sapere tutto di me, e quindi mette la bocca fra le mie cose.
c) Mia madre vuole sempre sapere tutto di me, e quindi mette le orecchie fra le mie cose.
d) Mia madre vuole sempre sapere tutto di me, e quindi mette i piedi fra le mie cose.

10.

a) Ieri sera sono tornata a casa da sola e ho avuto molta paura: poche luci, brutta gente. In questa città la notte c'è una cattiva aria.
b) Ieri sera sono tornata a casa da sola e ho avuto molta paura: poche luci, brutta gente. In questa città la notte c'è un brutto clima.
c) Ieri sera sono tornata a casa da sola e ho avuto molta paura: poche luci, brutta gente. In questa città la notte c'è una brutta aria.
d) Ieri sera sono tornata a casa da sola e ho avuto molta paura: poche luci, brutta gente. In questa città la notte c'è un cattivo clima.

L2 to L1 translation

Prova 3. Segna la traduzione più giusta (______/10)

1. Luca dice sempre di essere il più forte a tennis. E invece oggi ha perso la faccia con un bambino.
   a) Luca丢脸了。
   b) Luca和另外一个孩子玩了。
   c) 我记不住Luca的脸。

2. Sono andata a casa della mia amica Chiara, ma c'era una brutta aria, forse aveva litigato con i suoi genitori.
   a) 在Chiara的家有恶臭。
   b) 在Chiara的家有不好的气氛。
   c) 在 Chiara 的家很热。

3. Chiara non mi piace: quando viene a casa mia, mette il naso dappertutto.
   a) Chiara 感冒了。
   b) Chiara闻整个房子。
   c) Chiara环顾四周所有的房间。

4. Domenica scorsa a casa ho avuto una doccia gelata: il gatto ha mangiato il mio pesce rosso!
   a) 在我洗澡的时候, 猫把我的 金鱼吃了。
   b) 在我家没有热水了。
   c) 我回家的时候， 有一个不好的惊喜。

5. Non dovevi comprare questo cappello: costa un occhio!
   a) 帽子很贵。
   b) 帽子很丑。
   c) 帽子不适合你。

6. Il figlio di Carlo e Giulia è una testa calda e loro sono molto preoccupati.
   a) Carlo 和 Giulia的男儿很浮躁。
   b) Carlo 和 Giulia 的男儿很聪明。
   c) Carlo 和 Giulia 的男儿一直感觉很热。

7. In ufficio ho incontrato Luca e il suo braccio destro.
   a) 我察觉Luca的手臂。
   b) 我遇到了Luca和帮助他的人。
   c) Luca很强。

8. Ieri ho fatto una gita fuori città. Mi sembrava di toccare il cielo!
   a) 昨天我爬了很高的山。
   b) 昨天我为了短期旅行很高兴。
   c) 昨天我坐了飞机。

9. Ieri a cena abbiamo parlato solo di aria fritta.
   a) 昨天吃晚饭的时候我谈了做饭。
   b) 昨天晚饭我们吃了油炸食品。
   c) 昨天吃晚饭的时候我谈了不重要的事情。

10. Ieri credevo di poter guidare la moto di Marco, ma stamattina ho aperto gli occhi: è troppo difficile per me!
    a) 今天早晨我醒来的时候，我不会骑摩托车。
    b) 今天早上我觉得我不会骑摩托车。
    c) 今天我很早醒来。

Appendix E. Ethics Committee Letter

University of Portsmouth
Faculty of Humanities and Social Sciences

Wishing you every success in your research

Chair

Dr Jane Winstone

Email: ethics-fhss@port.ac.uk

Annexes

A - Documents reviewed

B - After ethical review

ANNEX A - Documents reviewed

The documents ethically reviewed for this application

Document	*Version*	*Date*
Application Form	2	13 Oct 2017
Participant Information sheets Pre-participation information sheet Post participation debriefing	2	13 Oct 2017
Consent Forms Pre-participation consent form Post debriefing consent form	2	13 Oct 2017
Supervisor Email Confirming Application	1	15 Sept 2017
Evidence from external organisation showing support	1	15 Sept 2017
Research proposal reviewed by supervisors	1	15 Sept 2017

ANNEX B - After ethical review

1. This Annex sets out important guidance for those with a favourable opinion from a University of Portsmouth Ethics Committee. Please read the guidance carefully. A failure to follow the guidance could lead to the committee reviewing and possibly revoking its opinion on the research.

2. It is assumed that the work will commence within 1 year of the date of the favourable ethical opinion or the start date stated in the application, whichever is the latest.

3. The work must not commence until the researcher has obtained any necessary management permissions or approvals – this is particularly pertinent in cases of research hosted by external organisations. The appropriate head of department should be aware of a member of staff's plans.

4. If it is proposed to extend the duration of the study beyond that stated in the application, the Ethics Committee must be informed.

5. Any proposed substantial amendments must be submitted to the Ethics Committee for review. A substantial amendment is any amendment to the terms of the application for ethical review, or to the protocol or other supporting documentation approved by the Committee that is likely to affect to a significant degree:

(a) the safety or physical or mental integrity of participants
(b) the scientific value of the study
(c) the conduct or management of the study.

5.1 A substantial amendment should not be implemented until a favourable ethical opinion has been given by the Committee.

6. At the end of the work a final report should be submitted to the ethics committee. A template for this can be found on the University Ethics webpage.

7. Researchers are reminded of the University's commitments as stated in the Concordat to Support Research Integrity viz:

- maintaining the highest standards of rigour and integrity in all aspects of research
- ensuring that research is conducted according to appropriate ethical, legal and professional frameworks, obligations and standards
- supporting a research environment that is underpinned by a culture of integrity and based on good governance, best practice and support for the development of researchers
- using transparent, robust and fair processes to deal with allegations of research misconduct should they arise
- working together to strengthen the integrity of research and to reviewing progress regularly and openly.

8. In ensuring that it meets these commitments the University has adopted the UKRIO Code of Practice for Research. Any breach of this code may be considered as misconduct and may be investigated following the University Procedure for the Investigation of Allegations of Misconduct in Research. Researchers are advised to use the UKRIO checklist as a simple guide to integrity.

Appendix F. Form UPR16

FORM UPR16

Research Ethics Review Checklist

UNIVERSITY OF PORTSMOUTH

Please include this completed form as an appendix to your thesis (see the Research Degrees Operational Handbook for more information

Postgraduate Research Student (PGRS) Information		**Student ID:**	867873
PGRS Name:	Ilaria Borro		
Department:	School of Languages and Applied Linguistics	**First Supervisor:**	Nick Bertenshaw & Alessandro Benati (Joint first supervisor)
Start Date: (or progression date for Prof Doc students)	01/02/2017		
Study Mode and Route:	Part-time ☐ Full-time ☒	MPhil ☐ PhD ☒	MD ☐ Professional Doctorate ☐

Title of Thesis:	Enhanced incidental learning of formulaic sequences by Chinese learners of Italian.
Thesis Word Count: (excluding ancillary data)	62020

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES ☑ NO ☐
b) Have all contributions to knowledge been acknowledged?	YES ☑ NO ☐
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES ☑ NO ☐
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES ☑ NO ☐
e) Does your research comply with all legal, ethical, and contractual requirements?	YES ☑ NO ☐

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	17/18:02

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:

Signed *(PGRS)*:	[signature]	**Date:**	12/01/2021

UPR16 – April 2018