UNIVERSITAT DE BARCELONA

Learning formulaic sequences from captioned videos: the effect of genre, metalinguistic awareness and statistical learning ability

Natalia Moskvina

Doctoral Dissertation

Doctoral programme in Cognitive Science and Language

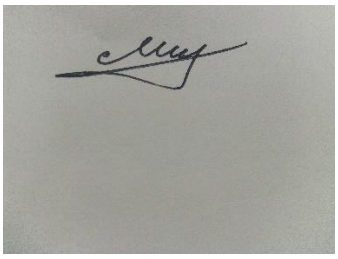
Learning formulaic sequences from captioned videos: the effect of genre, metalinguistic awareness and statistical learning ability.

Natalia Moskvina

Supervised by Dr. Roger Gilabert Guerrero

2023

Barcelona



Abstract

This doctoral dissertation examines incidental learning of formulaic sequences from exposure to L2 captioned videos of different genres and explores learner-related factors that might affect this process, such as metalinguistic awareness, statistical learning ability, prior knowledge of the items, and age.

To this end, data was collected from 114 Catalan/Spanish learners of English during two experimental sessions. During the first session, participants' metalinguistic awareness, statistical learning ability and vocabulary size were evaluated, and they were pre-tested on the knowledge of target formulaic sequences. Based on the scores from this session and learners' background characteristics, they were then assigned to one of the four treatment groups in a way that ensured the comparability of the groups. During the second session, each of the treatment groups watched approximately 30 minutes of captioned videos (in English) from one of the 4 genres: sitcom, TED talks, drama, or action, with the viewing immediately followed by the post-test. All genre groups were exposed to the same 6 target pragmatic formulaic sequences, and each item appeared twice within a genre.

The results confirm the overall beneficial effect of exposure to L2 captioned videos on acquisition of formulaic sequences. However, not all genres equally promoted learning, with the pre- post-test gains not being significant for the action group. Additionally, there were interactions between genre and age as well as between genre and vocabulary size. While sitcom and drama seemed to be suitable for learners of various ages and vocabulary sizes, TED talks were more advantageous for older participants and those with larger vocabulary knowledge. The opposite was the case for the action group, where younger participants and those with less vocabulary knowledge showed higher gains after the treatment. This allows

us to suggest that genres vary in the learning conditions they provide, and these differences can affect learning of formulaic sequences.

From learner-related factors, metalinguistic awareness and prior knowledge of the target items emerged as significant predictors of learning gains, with higher metalinguistic awareness having a faciliatory effect on learning and higher pre-test scores leading to lower gains. Individual differences in participants' statistical learning ability did not have any impact on the learning outcomes. It is therefore proposed that L2 leaners rely more on explicit learning mechanisms than on implicit ones, and it is important to be aware of formulaic language and its main aspects in order to acquire formulaic sequences incidentally. Besides, partial knowledge of formulaic sequences could be a hindering factor in that process.

Our findings also suggest that metalinguistic awareness is not a simple by-product of proficiency and can be examined at the level of individual differences. Learners' awareness varied across formulaic aspects and types as well, showing that they are not fully aware of the full spectrum of formulaic sequences and their defining features.

To conclude, this doctoral dissertation explores incidental learning of formulaic sequences from authentic and engaging input and the factors that might influence this process. It therefore advances our general understanding of formulaic language, its nature, and its acquisition.

Resum

Aquesta tesi doctoral examina l'aprenentatge incidental d'expressions idiomàtiques a partir de l'exposició a vídeos amb subtítols L2 de diferents gèneres i explora factors relacionats amb l'alumne que podrien afectar aquest procés, com ara la consciència metalingüística, la capacitat d'aprenentatge estadístic, el coneixement previ dels elements i l'edat.

Amb aquest objectiu, es van recollir dades de 114 alumnes d'anglès catalans/espanyols durant dues sessions experimentals. Durant la primera sessió, es va avaluar la consciència metalingüística dels participants, la capacitat d'aprenentatge estadístic i la mida de vocabulari, i se'ls va sotmetre a un pretest del coneixement de les expressions idiomàtiques objectiu. A partir de les puntuacions obtingudes en aquesta sessió i de les característiques de fons de l'alumnat, se'ls va assignar a un dels quatre grups d'intervenció de manera que es garantís la comparabilitat dels grups. Durant la segona sessió, cadascun dels grups d'intervenció va mirar aproximadament 30 minuts de vídeos subtitulats, en anglès, d'un dels 4 gèneres: sitcom, TED talks, drama o acció. Tot seguit, els alumnes van realitzar el posttest. Cada grup de cada gènere va ser exposat a les mateixes 6 seqüències d'expressions idiomàtiques objectiu, i cada element apareixia dues vegades dins d'un gènere.

Els resultats confirmen l'efecte beneficiós en general de l'exposició a vídeos amb subtítols L2 sobre els guanys idiomàtics. Tanmateix, no tots els gèneres van propiciar l'aprenentatge per igual, i el vocabulari adquirit entre el pretest i el posttest no va ser significatiu per al grup assignat al gènere d'acció. A més, hi va haver interaccions entre gènere i edat, així com entre gènere i mida de vocabulari. Si bé la sitcom i el drama semblaven adequats per a aprenents de diferents edats i mides de vocabulari, les TED talks eren més avantatjoses per als participants més grans i per a aquells amb més coneixement de vocabulari. En contrast, al

grup d'acció van ser els participants més joves i els que tenien menys coneixements de vocabulari els que van mostrar una adquisició major després de la intervenció. Això ens permet suggerir que els gèneres varien en les condicions d'aprenentatge que ofereixen, i aquestes diferències poden afectar l'aprenentatge d'expressions idiomàtiques.

A partir dels factors relacionats amb l'alumne, la consciència metalingüística i el coneixement previ dels elements objectiu van sorgir com a predictors significatius dels guanys d'aprenentatge, on la consciència metalingüística més alta té un efecte facilitador sobre l'aprenentatge i puntuacions més altes als pretests, la qual cosa condueix a guanys menors. Les diferències individuals en la capacitat d'aprenentatge estadístic dels participants no van tenir cap impacte en els resultats d'aprenentatge. Per tant, es proposa que els alumnes de L2 es basen més en mecanismes d'aprenentatge explícits que no pas implícits, i que és important conèixer fórmula el llenguatge idiomàtic i els seus principals aspectes per tal d'adquirir expressions idiomàtiques de manera incidental. A més, el coneixement parcial de les expressions idiomàtiques podria ser un factor obstaculitzador en aquest procés.

Els nostres resultats també suggereixen que la consciència metalingüística no és un simple subproducte de la competència lingüística i que es pot examinar a nivell de diferències individuals. La consciència dels alumnes també variava en funció dels aspectes idiomàtics i dels tipus d'expressions, la qual cosa posa de manifest que no són plenament conscients de l'espectre complet d'expressions idiomàtiques i les seves característiques definitòries.

Per concloure, aquesta tesi doctoral explora l'aprenentatge incidental d'expressions idiomàtiques a partir de contingut autèntic i atractiu i els factors que poden influir en aquest procés. D'aquesta forma, millora la nostra comprensió general del llenguatge idiomàtic, la seva naturalesa i la seva adquisició.

Acknowledgements

This doctoral dissertation would not be possible without a number of people who kindly helped me though the process.

First of all, I would like to thank my patient supervisor, Dr. Roger Gilabert. I am immensely grateful for all the great discussions that we had and all the support that you provided. Thank you for never rejecting my ever so unorthodox ideas and showing me that there is a place for humanity in academia.

The unique conditions that predoctoral researchers enjoy at the GRAL research group cannot be overestimated. I am thankful to Dr. Carme Muñoz and other senior members for always treating PhD students as colleagues and taking active interest in our work.

I was incredibly lucky in my research stays. Dr. Anna Siyanova-Chanturia hosted me at the Victoria University of Wellington during the challenging times of the pandemic onset, and kindly found time to listen to my ideas in the middle of a worldwide crisis.

My visiting time at Radboud University Nijmegen, although short, was immensely fruitful. My warmest gratitude goes to Dr. Ferdy Hubers and Dr. Catia Cucchiarini for inviting me to share my research, taking time to discuss random slopes and of course, introducing me to pepernoten.

Finally, my sincerest gratitude to my amazing friends and family, who have been asking me the whole day today if I have submitted yet. Now yes!

Table of contents

List of tables

List of figures

Introduction

Despite what we might believe about the uniqueness and creativity of the discourse that we produce, our speech to a large extent consists of recurring combinations of words, or formulaic sequences. According to Erman and Warren (2000), formulaic sequences correspond to approximately 52% of written and 58% of spoken discourse, which suggests that first language speakers and writers rely extensively on formulaic language.

Wray (2012) puts forward an interesting explanation of this extensive presence of formulaic sequences in first language use, highlighting its relevance for successful social interaction ("*I am like you because I talk like you, so you will want to help me*": Wray, 2012, p.232) and avoiding misunderstanding. She suggests that the widespread use of conventionalized sequences stems from our instinct to belong to and be recognized as a member of a community.

Since formulaic language seems to be an essential part of native language competence, it is important for second and foreign language learners to acquire this language aspect as well. However, they find it particularly challenging, overusing the items they are familiar with and underusing others required by context (see Paquot & Granger, 2012, for a review). One of the reasons for this difficulty in formulaic acquisition lies in the fact that input available to language learners is limited and does not always reflect real-life native use (e g., Northbrook & Conklin, 2018). In order to address this issue, researchers are increasingly turning to TV programmes in the target language for authentic audio-visual input. The potential of this type of input to promote learning of formulaic sequences has been demonstrated at a more theoretical corpus level (Lin, 2014) as well as at the experimental one (Puimège & Peters, 2019a,2020; Majuddin et al., 2021, Vu et al., 2023). However, providing learners with rich

input is not always sufficient and an effort should be made to ensure that they can profit from such exposure. Two types of factors are typically considered in an attempt to create optimal learning experience: context-related and learner-related ones. Genre of audio-visual materials is selected here as a potential context-related factor explaining variations in learning conditions provided by different L2 videos (Webb & Rodgers, 2009a, 2009b; Moskvina, 2017; Gilabert et al., 2018; Suarez et al., 2021). These variations could be crucial for formulaic sequences since they rely quite heavily on the context.

As for learner-related factors, since formulaic sequences are often not recognized as such in the input (Bishop, 2004; Martinez & Murphy, 2011; Guz, 2014, Kim, 2016), we propose metalinguistic awareness and statistical learning ability as two possible mechanisms underlying this process. The role of background factors, such as age, vocabulary size, and prior knowledge of the target items is discussed as well.

Thus, the objectives of the present doctoral dissertation are to further explore the acquisition of formulaic sequences from audio-visual input and to examine the context- and learner related factors that have not been brought into focus before. These factors might shed light on the optimal ways to create a multimedia instruction as well as deepen our understanding of how different types of learners can profit from this instruction.

This dissertation is based on one principal study and consists of four main chapters: literature review (Chapter 1), methodology (Chapter 2), results (Chapter 3), and discussion (Chapter 4). In the literature review chapter, the overview of relevant previous research is provided in order to contextualize the study and familiarize the reader with its key concepts, and the research agenda for the current project is specified. The methodology section describes the research design of the study and all its elements and is concluded with the preliminary

analysis of the collected data. The results chapter presents the statistical analysis carried out to answer the proposed research questions along with the findings of this dissertation. The final chapter contains the discussion and interpretation of the findings, limitations on their generalizability as well as general conclusions and pedagogical implications.

Wray (2012) illustrates research on formulaic language through the parable of the blind men and an elephant, where the men provide contradicting descriptions of the animal since they have access to its different parts. Complexity of the phenomenon in question and the variety of approaches to its investigation are very accurately reflected in this comparison. This doctoral dissertation possibly provides a picture of elephant's left back leg, but we hope that it contributes to the collective effort to further our description and understanding of the formulaic elephant.

Chapter 1. Literature review

The present chapter provides the theoretical background for the linguistic phenomena covered in the dissertation as well as the summary of the relevant empirical research. It includes 3 main sections organized in the following way: first, the language aspect in focus is introduced. Further on, learner-related variables potentially affecting acquisition of the target language aspect are discussed. Finally, the selected learning context is described as well as the distinct learning conditions it offers, and the underresearched context-related factor is specified. The chapter is concluded by the identification of the research gap, determination of research objectives and the research questions formulation.

1.1. Formulaic language

1.1.1. Defining the phenomenon

In contrast with the previous exclusive focus on individual words, a growing number of vocabulary studies favor recurrent multiword units as essential items of the lexicon. These units are currently united in the literature under the term "formulaic language", however, they vary considerably across a range of aspects and there is little agreement over the terminology applied and the items that should be grouped within its scope. Wray (2002) lists over 60 terms used to describe the phenomenon in question, e.g., *set phrases, chunks, multiword items/units, fixed expressions, prefabricated routines and patterns*, etc. In this dissertation we will adopt *formulaic sequence/item* as the most general and neutral term, alternating it with *multiword unit/sequence*. Although some researchers draw a distinction between the two denominations, stating that *formulaic sequence* can refer to syllable or morpheme strings as well as those of words, while *multiword unit* exclusively describes strings consisting of two or more words, the application of this distinction is quite narrow and is usually motivated by a specific research context. Similarly, for the purposes of this dissertation we will not focus on individual words as instances of formulaic language, despite recognizing that it is the case in some studies (e.g., Wray, 2002).

Since clear identification of the concepts plays an important role in setting the research trajectory, we start by proposing the following definition of formulaic sequences, based on Siyanova-Chanturia and Pellicer-Sánchez (2018): *a variety of conventionalized sequences that are recognized by the members of a language community as a common and the most natural way of expressing a certain meaning.* We believe that it reflects the most important aspects of the studied phenomenon: widespread nature and familiarity for a typical speaker

of a language as well as the arbitrary connection between form and meaning based solely on convention. For example, there is nothing wrong with the expression "*strong rain*" from the grammatical or semantic point of view, however, it is by far less common and natural for a native English speaker than "*heavy rain*".

The interpretation of the scope of formulaic language varies across the studies, with two main approaches being distinguished: usage-based and phraseological.

Usage-based approach (also known as distributional or frequency-based) relies on distributional data from corpora in the identification of formulaic sequences. (e.g., Sinclair, 1991; Biber, 2009; Wulff, 2018). Multiword combinations are recognized as formulaic if they occur in the language frequently enough, regardless of their idiomaticity. Thus, according to this approach, such units as "*have a green thumb*" and "*I thought that was*" would be assigned the same formulaic status.

No type-oriented classification of formulaic sequences has been advanced under the usage-based approach, as it primarily aims to reflect real-life language use rather than categorize it in theoretical terms (for an exception, see the functional classification of lexical bundles in Biber & Barbieri, 2007). Within this framework, if any not frequency-related description of formulaic items is provided, they are typically characterized through their grammatical attributes, such as length, parts of speech or construction type, with construction grammar being one of the research paradigms associated with this approach. Instead, a great deal of attention is devoted under this perspective to determining what could be considered "frequent enough". Besides raw frequency thresholds (e.g., 40 times per million words in Biber et al., 2004), association measures are commonly applied to identify formulaic items. Some of the widely used measures include mutual information and t-score.

Mutual information (MI) reflects association strength between two words within a sequence, typically a collocation. It is based on the ratio between the times the words appeared in the corpus as a part of a given unit and the times they individually occurred outside of it. The drawback of this measure is its skewness towards completely fixed sequences containing low frequency words, as their frequency of occurrence within a sequence would be naturally higher compared to that at the individual level.

This concern is addressed in the *t-score* measure, which reflects the degree of confidence in the association between two words within a sequence rather than its strength. It shows whether the words are combined together by pure chance or there is a statistically significant pattern in their co-occurrence. It could be challenging, however, to establish the reliable level of significance and compare it across corpora. Unlike mutual information, t-score favors highly frequent sequences, sometimes failing to distinguish between formulaic items and free combinations.[1]

A number of studies turned to distributional properties in order to explain processing and knowledge of formulaic sequences by native speakers and language learners (e.g., Ellis et al., 2008; Durrant & Schmitt, 2010; González Fernández & Schmitt, 2015; Nguyen & Webb, 2017). While general statistical properties certainly account for many aspects of formulaic use, the presence of type-specific factors is suggested as well. For example, Schmitt et al. (2004) explored the validity of corpus-driven formulaic sequences through an oral dictation task, where English speakers and learners had to repeat back the dictated chunks of text

[1] In this dissertation we use the term free combinations to refer to word combinations that are outside of the formulaic scope (e.g., *blue lamp*), unlike some other studies (e.g., Gyllstad & Wolter, 2016; Puimège & Peters, 2020) that apply it to the category that is called here *unrestricted collocations*. More information is provided below.

instead of writing it down. The text contained 25 formulaic candidates of various frequencies, extracted from several corpus-based lists. It was predicted that if presented with chunks of text that are long enough to push their working memory capacity to the limit, participants would be more fluent and accurate in repeating the sequences that are truly formulaic and stored as such in the mental lexicon, while struggling with the rest. Besides expected differences between native and non-native groups, the effect of sequence properties was reported. Their frequency did not affect participants' performance on the dictation task, while semantic and functional transparency (how easy it was to guess the meaning or function of the items) did. The pattern was the same for the native speakers as well as for the L2 learners, with the transparency effect being more pronounced for the latter group. The varied performance allowed the authors to infer that not all recurrent corpus-driven sequences are equally formulaic from the psycholinguistic point of view. It was also pointed out that corpus frequency was not sufficient to explain the processing differences across the sequences, thus emphasizing the importance of other formulaic aspects.

This finding is partially supported by Columbus (2010). The eye-tracking reading study showed processing advantage of idioms (*lost his marbles*) over restricted collocations (*played some part*) and lexical bundles (*would you like to*), corroborating the psycholinguistic validity of formulaic types. However, unlike Schmitt et al. (2004) the frequency effect was reported as well. The author concludes that the formulaic types do not exist independently but rather are defined through a combination of various formulaic features.

The features that might be helpful in differentiating between formulaic types are further explored by Columbus (2013). The results show that English idioms, restricted collocations, and lexical bundles contrast in corpus frequency as well as semantic transparency and

familiarity judged by native speakers. The three discussed types fit naturally on the scale from the most frequent, familiar and transparent (lexical bundles) to the least frequent, familiar, and transparent (idioms), with restricted collocations being placed in the middle.

Finally, an eye-tracking study by Carrol and Conklin (2020) examined relative contribution of general and type-specific features to the processing advantage of three formulaic types: idioms, binomials (e.g., *salt and pepper, pack and choose*), and collocations. The considered features included the following general distributional measures: adjusted phrase frequency, cloze probability (describes how easy it is for a native speaker to guess the final word in a sequence) and transitional probability (reflects the corpus-based likelihood of one word occurring after the other). Type-specific predictors involved semantic association (semantic relationship between the constituent parts for collocations and binomials), familiarity (rated by native speakers, used for idioms), decomposability (rated by native speakers, defines the degree to which individual words contributed to the figurative meaning of the idioms), reversibility ratio (corpus-based ratio of forward and backward occurrences for binomials), and mutual information (for collocations). Besides the overall processing advantage of formulaic items vs. non/formulaic ones, the results show that the set of factors best explaining the observed reading patterns varied for different types. For idioms, the most consistent facilitatory effect on processing was that of phrasal frequency, with decomposability and familiarity affecting certain aspects of participants' eye behavior as well, negatively and positively correspondingly. Phrasal frequency and cloze probability best explained reading patterns for binomials, while for collocations no effect of frequency was present, and cloze probability along with mutual information facilitated processing instead. Although most of the major predictors belong to the same class of distributional properties, it was demonstrated

that they best explain processing of formulaic sequences when their statistics is tuned in to the peculiarities of different formulaic types. For example, in line with previous research (Ellis et al., 2008), mutual information emerged as a more efficient measure of collocation strength than frequency.

The observed patterns allow us to suggest that despite the undoubted importance of distributional properties, they alone cannot account for the variability in the processing and learning of formulaic language. It is clear that we are not achieving full understanding of the phenomenon in question merely by registering what is frequent in language use, and the analysis of the features related to its non-distributional aspects might deepen our insight into the matter.

Various classifications of formulaic sequences have been proposed under **phraseological approach** (adhered to, for instance, by the following authors: Allerton et al., 2004; Nesselhauf, 2005; Burger et al., 2007)**.** This approach determines formulaic sequences in terms of their semantic, structural, and functional properties, rather than statistical ones. At its heart traditionally lie idioms or idiom-like, fully non-compositional expressions as the prototypical formulaic item. (Non-)compositionality, defined by the relationship between the overall meaning of the formulaic expression and individual meanings of its constituents, is often used as a primary criterion for formulaic selection and classification. Other common criteria include internal structure (e.g., verb + noun,), extent (phrase or sentence level), degree of syntactic and lexical flexibility (ability to allow for structural or lexical variation), discourse function (e.g., clarifying, turn shifting, connecting) (Granger & Paquot, 2008). Different phraseological taxonomies vary in the choice of formulaic dimensions used for creating categories as well as in their scope (e.g., lexical and grammatical collocations vs. a

single category of collocations), resulting in less homogeneity in the terminology for this research tradition.

As we believe that both usage-based and phraseological approaches reflect crucial aspects of the studied phenomenon, in the present dissertation we use a combination of the two, discussing phraseological features as well as distributional ones.

The following classification of formulaic sequences is adopted, based on the one described in Hubers (2020): idioms, collocations, phrasal verbs, binomials, pragmatic sequences (speech formulas), and lexical bundles.

Idioms or *idiomatic expressions* are formulaic sequences that have a recognizable figurative component in their meaning. Examples of idioms are: *kick the bucket, to have green fingers*, *once in a blue moon*. While other types might exhibit certain degree of figurativeness as well, it is not their defining feature, unlike for idioms. Idioms can differ greatly in structure and grammatical characteristics, but they invariably carry figurative meaning of some strength and therefore are never fully compositional. It is noteworthy that the relationship between the figurative and literal interpretation of an idiom is not that straightforward, with considerable fluctuation in transparency (the extent to which the figurative meaning can be derived from the literal one), semantical decomposability (the extent to which individual words map onto the figurative idiom meaning), and imageability (how easily an idiom can be visualized) (Hubers, 2020).

Collocations are usage-determined frequently co-occurring pairs of words. They are often described in terms of collocate-node relationship, i.e., the parts of speech involved (verb + noun, adjective + noun, verb + adverb, etc.). Traditionally, phraseology focuses on the pairs

where the choice of the collocate is limited or restricted by conventional use (restricted collocations), while corpus studies bring frequent unrestricted collocations into focus as well. Both types are considered to belong to the same collocation category here, and are differentiated from free combinations (lexically unrestricted and not frequent). Examples of restricted collocations are: *make a mistake* (not do a mistake), *plastic surgery* (not plastic operation), *deep sleep* (not profound sleep). Examples of unrestricted collocations are: *great idea* (also could be *nice/good/excellent/brilliant idea*), eat lunch (also could be *have/grab lunch*). Examples of free combinations are: *read a line, want a car, yellow cup*. It has to be noted that the line between the categories is a fine one, so it is interpreted as more of a direction rather than a clear cutting point.

Following Heid (2002) and in contrast with Granger and Paquot (2008) we do not subdivide this category into lexical *(heavy rain, face a problem*) and grammatical collocations (*aim at, interested in*). We believe that grammatical collocations are better described in terms of verb valency and therefore belong to syntax rather than to phraseology. As a result, when we discuss collocations in this study we are referring exclusively to the lexical ones.

Meaning shift units (MSUs) are acknowledged in the present dissertation as a special collocation type that involves meaning shifts, i.e., changes in the meaning of individual words within a collocation that happen due to their co-occurrence (e.g., *deep trouble* vs. *deep see*). This interpretation of MSU differs from the initial understanding of this term by Sinclair (2007a, 2007b), who proposed MSUs as a more accurate denomination for collocations, with co-selection induced meaning shifts, however subtle, used as a criterion for defining this category. Our view is better explained by Vetchinnikova's (2014) model, which includes

MSU as the last stage of delexicalization process[2] within a collocation pair and as the first step towards forming an idiomatic expression.

Phrasal verbs are verb + particle(s) combinations that convey a single meaning and thus function as single words (and could be replaced by such). Examples of phrasal verbs are: *work out, give in, put up with*. These combinations may vary from literal ones (e.g., *move in*) to those with more figurative meaning (e.g., *let down*), and are often polysemous (e.g., *take in* can mean deceive, understand, or provide a place to stay). It is one of the most challenging types for second language learners (Liao & Fukuya, 2004; Garnier & Schmitt, 2016; Peters, 2016).

Binomials are recurrent pairs of words that belong to the same part of speech and are linked by a conjunction (usually *and, or*) or a preposition (e.g., *by*). Examples of binomials are: *life and death, black and white, more or less, step by step*. They are normally fully compositional, i.e. their meaning is directly derived from the two parts of the pair. Despite the structural peculiarities, the most important feature of binomials is strong preference for a specific word order, which is based on convention rather than on any semantic criteria. There is no explicit linguistic restriction preventing native English speakers from saying "*white and black*", however, this variant will appear 974 times in Corpus of Contemporary American English against 8450 times for the preferred word order. As well as the other type-specific features, reversibility of binomials varies for different instances of the formulaic type in question, with some items showing stronger preference in word order than others.

[2] A process where a word within a collocation loses its original meaning.

A special subgroup of binomials is comprised by *irreversible binomials*, those that have the highest degree of fixedness and never appear in the corpus in the alternative order. Examples of irreversible binominals are: *hit and run, by and large, macaroni and cheese*. Some of these binomials are conditioned by the logical sequence of events or objects (e.g., *catch and release*; you cannot release something before catching it first). In addition, they often acquire figurative meaning and lose their compositionality, thus functioning very similarly to idioms (Arcara et al., 2012; Carrol & Conklin, 2020), for instance, *hit and run* describes a specific type of a road accident rather than any other situation involving these actions.

Lexical bundles are highly frequent recurrent word sequences that normally consist of 3-6 words and are corpus-driven (extracted from the corpus based on a set frequency threshold). Since the main criterion is frequency of occurrence and neither idiomacity nor structure are taken into account, this category tends to be highly heterogeneous. Examples of lexical bundles are: *a little bit more, in any case, to begin with, something like that, is going to be.*

Pragmatic formulaic sequence, also known as *speech formulas, pragmatic routines, conventional expressions, or situation-bound utterances (SBUs)* are formulaic items that perform a certain pragmatic function and are closely linked to a particular social context (e.g., greeting, accepting apology, clarifying the message, confirming understanding, etc.) Examples of pragmatic formulaic sequences are: *don't get me wrong, you're welcome, no problem, how are you?* These items vary in compositionality, with such sequences as *that'd be great* or *I'm just looking* being fully compositional, and others like *come again?* or *tell me about it!* having more idiomatic, non-compositional meaning. Difficulty of acquiring pragmatic formulaic sequences lies in the fact that learners have to achieve not only pragmalinguistically correct performance, i.e., use the correct form of the sequence (*don't*

get me wrong and not *don't put me wrong*) but sociopragmatic appropriateness as well, i.e., use the sequence in the appropriate context (*don't get me wrong* is used to avoid misunderstanding and conflict when a potentially negative or upsetting message is conveyed).

Some researchers, for example Kecskes (2010) identify pragmatic formulaic sequences as a broad class of conversational routines and further categorize it into the subgroups of situation-bound utterances (linked to a particular situation: meeting a new person, getting on a plane, e.g., *how do you do, happy birthday, welcome aboard*), speech formulas (linked to a particular function: agreeing, confirming understanding, e.g., *you know, I see*), and discourse markers (*nevertheless, in contrast*). We will not follow this classification here, using the terms pragmatic formulaic sequences and speech formulas interchangeably, and not considering discourse markers as a part of the category.

Certain formulaic categories are not recognized in the present dissertation despite being previously discussed in the literature. For example, compounds are not included here as a separate formulaic type. There is no unified agreement over the phraseological status of these items, with some researchers arguing that they belong to morphology instead (Moon, 1998) Relying on the graphic criterion rather than the semantic one, we view one-word compounds as well as the hyphenated ones as single words (e.g., *bedroom, mother-in-law*), and the ones spelled as two words as collocations (e.g., *living room, coffee table*). Although variability in prescribed spelling makes this categorization somewhat unstable, we believe that conjoint and hyphenated compounds exhibit a clear visible connection between the two constituents,

a feature that formulaic sequences typically do not possess[3]. Further discussion of the place of compounds on a phraseological spectrum, while engaging in its own right, is left outside of our research scope.

Similes and proverbs were excluded from the classification as well. Figurative similes are regarded here as idioms (e.g., *cute as a button, as busy as a bee*), while more literal ones are not considered formulaic (e.g., *blue as the sky*). Proverbs, although certainly displaying a range of formulaic features, are a very special class of items. While most formulaic sequences are conventionalized ways to express common meanings, proverbs can be viewed as conventionalized ways to express conventionalized meanings. Thus, we recognize its formulaic potential but do not include it in the present discussion due to its unique status.

The classification described above is primarily based on the phraseological categories, however, it acknowledges the importance of distributional properties by including the category of lexical bundles and using corpus data to describe other types (e.g., binomials and collocations).

While it is important to define all the formulaic types clearly as they are repeatedly referred to throughout the dissertation, the learning focus of our research is limited to pragmatic formulaic sequences. These formulaic items are traditionally reviewed in purely pragmatic terms and are rarely included in the general discussion on the nature of formulaic language and its properties. Intervention studies on learning of pragmatic formulaic sequences are also relatively scarce (Bardovi-Harlig, 2012).

[3] This refers to written speech, as previous studies did show phonological links between constituents of a formulaic sequence in spoken language (e.g. Lin, 2018)

After explaining our understanding of formulaic language and its categories, and identifying the learning focus of our research, we continue by discussing the importance of formulas for native speakers and second language learners.

1.1.2. Formulaic language in L1: importance of the phenomenon.

Formulaic sequences, understood in broad terms, are an important part of a language, and native speakers rely on them quite heavily (Sorhus, 1977; Pollio et al., 1977; Pawley & Syder, 1983; Sinclair, 1991; Erman & Warren, 2000). Sorhus (1977) reported that based on a Canadian corpus of spoken English language, an average native speaker uses one formulaic (she refers to it as "fixed") expression for every five individual words. Similarly, Erman and Warren (2000) showed that multiword units comprised about 52% of written and 58% of spoken discourse, thus confirming the largely formulaic nature of language.

Sinclair (1991) suggested that language production is governed by the alternation between the *idiom principle* (using prefabricated word strings) and the *open choice principle* (generating new combinations word-for-word), where the idiom principle is applied first and the open choice is resorted to when it is deemed insufficient. Although this theory evokes the controversial issue of holistic storage of formulaic items, its contribution to shifting the focus of vocabulary research from single words to longer sequences cannot be underestimated (Siyanova-Chanturia & Martinez, 2015).

1.1.3. Formulaic language in SLA: challenges.

In the field of second language acquisition, the significant contribution of formulaic competence to the native-like performance has been recognized as well. For example, Boers et al., (2006) reported a positive correlation between the use of formulaic sequences and

perceived oral proficiency. The results are further supported by Crossley et al. (2015), where collocation accuracy emerged as the strongest predictor of overall lexical proficiency in both oral and written production.

However, despite the evidence suggesting their importance and in contrast with native speakers' use, language learners' repertoire of formulas is usually poor and unbalanced (e.g., Kecskés, 2007). In fact, formulaic competence is one of the most challenging aspects for language learners. Besides producing incorrect formulaic sequences, they tend to overuse the items they are familiar with and underuse others, favored by context or convention (see Paquot & Granger, 2012, for a review), or to avoid formulas altogether in an attempt not to sound too cliché (Kecskés, 2000). The difficulties in producing formulaic language persist even at the advanced proficiency levels (Nesselhauf, 2005; Ellis et al., 2008; Thewissen, 2008).

One of the difficulties that contributes to learners' poor formulaic performance emerges at the very first stage of acquisition: formulaic sequences are frequently not recognized as such in the input. For instance, in Bishop (2004) participants' glossing behavior was observed during a computer-based reading comprehension task. The results showed that unenhanced unknown formulaic sequences were looked up significantly less often than unknown words, while both were needed to answer the comprehension questions. The author explains it by the fact that unlike word boundaries, clearly marked by spacing, the boundaries of new formulaic sequences are opaque for language learners. Typographic enhancement is suggested as the pedagogic tool to address this issue. Various formulaic types were included in the study, with phrasal verbs and idioms being most prominently represented.

In the same line of thought, Martinez and Murphy (2011) investigated the effect of idiomatic expressions on actual and self-reported comprehension. One hundred-one Brazilian learners of English (intermediate to advanced) were presented with 4 pairs of matching texts, where the two texts within a pair consisted of exactly the same highly frequent words. However, in one of the texts some of the words appeared as a part of an idiomatic expression, and in the other the very same items were used as single words. The evidence suggested that learners' actual comprehension was significantly lower for the texts where the words comprised a formulaic sequence than for those where they were used individually, and participants were significantly more inclined to overestimate their understanding of the texts containing idiomatic expressions than of the ones that did not contain any[4]. The results were interpreted in terms of Laufer's (1989a) concept of *deceptive transparency*: it was suggested that participants were "deceived" by the apparent familiarity of the single words in the idioms, which led them to misunderstanding the texts without realizing it. It was additionally pointed out that language learners seemed to vary in their perception of idiomaticity. Some of them showed no differences in reading times between the two types of texts, which, combined with their overestimated self-reported comprehension of the idiomatic texts allowed the authors to suggest that they were not aware of the formulaic nature of those texts.

Learners' ability to notice unknown idioms in the input was further explored by Kim (2016). The study applied a series of metalinguistic tasks to investigate the recognition of unfamiliar idioms by L2 learners of English (reading + identifying, reading + defining). The evidence suggested that unfamiliar idioms consisting of familiar words were recognized as such

[4] The study focused mainly on idioms, so all non-idiomatic sequences (e.g., lexical bundles, collocations) were not brought into consideration.

significantly less often than unfamiliar words, and the definitions of reportedly familiar idioms were significantly less acceptable than those of familiar words, corroborating the results from Martinez and Murphy (2011). Besides, the survey that participants took before the experiment showed that nearly half of the sample had none to very limited idea of what an idiom was and were not able to provide any examples, and the rest displayed a very narrow understanding of the concept, focusing on highly figurative idioms (e.g., *it's raining cats and dogs*) or items that are usually classified as proverbs (e.g., *all work and no play makes Jack a dull boy*). Summing up the results, the author concludes that language learners should be made aware of various types of multiword sequences present in the input to be able to acquire them incidentally.

These findings have an important implication for learning. Since language learners seem to be systematically unaware of the gap in their formulaic knowledge, it is doubtful that they will make any effort to bridge it. The Noticing Hypothesis (Schmidt, 1990) indicates that language items that are not noticed in the input, are unlikely to be acquired, which suggests poor prognosis for the learning of formulaic language so far.

To sum up what has been discussed so far, formulaic sequences are an important part of native speakers' vocabulary and it is therefore highly advisable for L2 learners to acquire them. However, language learners find these items very challenging, with problems starting at the very first stage of encountering new sequences and recognizing them. The evidence demonstrates a generally low level of noticing of formulaic language by L2 learners. As this can potentially result in poor learning outcomes, ways to facilitate noticing should be explored along with the factors affecting it. It is suggested that a closer examination of learners' awareness of formulaic language might be helpful in understanding their idea of

what comprises formulaic language as well as in determining how this idea can be modified through further instruction in order to promote formulaic competence. Schmidt (1990, 2001) distinguishes *awareness at the level of noticing* (conscious registering of instances of language use) vs *awareness at the level of understanding* (generalizing across the noticed instances of language use to form an explicit metalinguistic rule). Here, instead of this hierarchical structure that assumes varied degrees of processing depth, metalinguistic awareness is proposed as a general term that involves noticing and generalizing as interconnected processes that do not necessarily result in forming an explicit rule. The theoretical concept of metalinguistic awareness and its operationalization for formulaic language are discussed in more detail below.

1.2. Learner-related factors

1.2.1. Metalinguistic awareness: general concept

While language knowledge is knowledge of its items and ability to use them to communicate, metalinguistic knowledge or *metalinguistic awareness*[5] is knowledge about the language as a system, i.e., about its structure, constituting elements, and their functions.

Certain evidence has been obtained to support the claims about a facilitating role of metalinguistic awareness in SLA (e.g. Herrarte, 1998; Roehr, 2008; Woll, 2018). However, the results are not conclusive due to the complexity of the construct and significant differences in its operationalization.

[5] The two concepts are used here interchangeably.

Metalinguistic awareness is most typically viewed as explicit grammar knowledge and evaluated through error corrections or grammaticality judgements (e.g., Alderson et al., 1997; Elder & Manwaring, 2004), however, a number of researchers voiced their concern over this narrow approach (e.g., Roehr, 2008; Woll, 2018).

Woll (2018) defines metalinguistic awareness as a mental activity that involves deliberate attention allocation to language elements in the input. We find this definition particularly appealing as it highlights the importance of "knowing what you are looking for" at the initial stages of learning. This ability could be highly beneficial for recognizing formulaic sequences as special units in the input and selecting the important aspects of their occurrence for further analysis and assimilation: do these words always appear in this particular sequence? Can this sequence have any variants? Why doesn't this word in this sequence have the meaning I am familiar with?

The features of formulaicity that could potentially constitute metalinguistic awareness of formulaic language are further reviewed in the following section. In plain words, we will try to establish what learners need to know about formulaic language in order to efficiently notice it in the input, increasing the probability of its acquisition.

1.2.2. Metalinguistic awareness of formulaicity

Although the concept in question has not been consistently brought into focus in the context of formulaic language, certain potentially related aspects have been touched upon.

For example, a number of studies resorted to awareness-raising or awareness-based activities as a possible method to facilitate learning of formulaic sequences, with mixed results reported (e.g., Boers et al., 2006; Bardovi-Harlig & Vellenga, 2012; Wang et al., 2022). While most

of the activities did result in learners' increased awareness of formulaic presence in the input, it did not necessarily lead to increased formulaic competence. One suggested explanation for this finding is insufficient and inconsistent exposure to formulaic sequences (see Boers & Lindstromberg, 2012, for a review).

Learners' sensitivity to certain formulaic aspects have been demonstrated as well. For instance, Fioravanti et al. (2021) showed that although L2 Italian learners' intuitions about three categories varying in lexical fixedness (free combinations, collocations, idioms) differed from those of L1 speakers, they perceived idioms as the most inflexible class, similarly to native speakers.

In the same line of research, Hubers et al. (2020) measured Dutch learners' and native speakers' intuitions of idiom properties (imageability, transparency, familiarity) as well as their objective idiom knowledge. The results suggest that learners' intuitions of idiom properties are more accurate predictors of their actual idiom knowledge than native speakers' intuitions. This finding strengthens the connection between the way language learners perceive formulaic items and how well they know them.

Formulaic sequences are defined by a number of aspects, such as compositionality, fixedness, imageability, frequency, etc., and these aspects are potentially perceived by language learners, with a varying degree of accuracy. Therefore, we believe that metalinguistic awareness of fomulaicity should not be limited to the mere ability to notice formulaic items but rather should be viewed as a complex phenomenon incorporating at least some of these features. To this end we propose the following definition of the concept for the purposes of our study: *metalinguistic awareness of formulaicity is sensitivity to the essential properties of formulaic language.*

As the definition suggests, the next step is identifying what formulaic properties could be considered essential. Table 1 shows the features of formulaic sequences that are most commonly discussed in the literature.

Table 1. Formulaic features discussed in the literature

Feature	Associated formulaic types	Type of feature	Exclusively formulaic vs. general
Identifiability	All types	Common	General
Mutual information	Collocations	Statistical	Formulaic
t-score	Collocations	Statistical	Formulaic
Meaning shift	(Mainly) collocations	Meaning organization	Formulaic
Imageability	Figurative types	Meaning organization	Mostly formulaic
Transparency[6]	Figurative types	Meaning organization	Mostly formulaic
Compositionality	Most of the types	Meaning organization	Formulaic
Fixedness	Most of the types	Meaning organization	Formulaic
(Phrasal) frequency	All types	Statistical	General
Familiarity	All types	Statistical (subjective)	General
L1-L2 similarity	All types	Common	General
Perceptual (saliency)	All types	Common	General
Pragmatic function	Pragmatic formulas	Common	General
Figurativeness/literalness	Figurative types	Meaning organization	Mostly formulaic

The features in the table are classified according to the following criteria: formulaic types that they can be associated with (e.g., all the types can be described in terms of frequency, but pragmatic function is typically used to describe only pragmatic formulas), whether they are a general linguistic feature characterizing various language items (words, phonemes, etc.) or a special feature of formulaic items, and whether they belong to common, statistical, or meaning-organization properties of formulas.

[6] Used here in the meaning of paradigmatic transparency (Geeraerts, 1995)

While the common type includes the features that are not specific to formulaic language and therefore are not of much interest for our purposes (except for identifiability), the other two types are based on two important classes of formulaic properties: statistical regularities (statistical type) and peculiarities in meaning organization (meaning organization type). As was mentioned at the beginning of the literature review, formulaic items are often identified or described in distributional terms (phrasal frequency, mutual information, t-score). Another important aspect that characterizes most formulaic types (to a varying degree) is peculiarities in meaning organization. These peculiarities stem from the way meaning is mapped onto the form of the sequence, i.e., the relationship between individual meanings of constituting words and that of the whole expression (compositionality), certain meanings being fixed to particular forms not allowing any variability (fixedness), links between literal and figurative meaning components (imageability, transparency, figurativeness), changes in meaning due to co-occurrence (meaning shift).

It has to be acknowledged that there is certain variation in the way different researchers refer to some of these features. For example, the feature that here is called compositionality, is also known in certain contexts under the terms semantic decomposability (Titone et al., 2015), semantic analyzability (Gibbs, 1991), and syntagmatic transparency (Geeraerts, 1995). Compositionality was selected as it is frequently named as one of the main formulaic dimensions (e.g., Wray, 2012: Fioravanti et al., 2021), and the choice, as we see it, helps to avoid confusion between this feature and that of transparency.

Identifiability is a term that is proposed here to refer to the ability of formulaic sequences to be noticed (identified) in the input. It is similar to perceptual saliency, however, the latter is a more complex concept that potentially builds on other aspects of formulaicity and implies

that a described item contrast with the language elements surrounding it, while identifiability simply states if an item is consistently identified in the input or not, without any claims about why it happens.

In order to continue identifying what constitutes the essential properties of formulaic language, we need to take a closer look at two main classes of formulaic features, statistical and meaning organization ones. While statistical properties definitely belong to the core features of formulaic language, we cannot expect sensitivity to frequency and other distributional patterns to be efficiently evaluated through explicit metalinguistic awareness, as this sensitivity would involve other mechanisms of a more implicit cognitive nature. The cognitive ability of unconsciously detecting statistical patterns present in the input is commonly assessed in the literature as statistical learning ability. Therefore, statistical properties of formulaic language are recognized as essential but are excluded from our operationalization of metalinguistic awareness of formulaicity and learners' sensitivity to these properties is further discussed in the section on statistical learning ability below.

If we further examine the meaning organization properties, it becomes apparent that there is a considerable overlap between the features that describe the relationship of the overall meaning of a formulaic sequence and its various components, i.e., the features of compositionality, transparency, imageability, and figurativeness. Since these features seem to share an underlying principle, only one of them was selected to be included in the further operationalization of metalinguistic awareness. The feature of compositionality was chosen for that purpose, as it was viewed as the most general and representative, and could be used to describe a wide range of formulaic types, while the other three are closely connected to formulaic items that have a recognized figurative component.

Thus, the analysis of formulaic properties discussed in the literature allow us to define the following aspects as core components of metalinguistic awareness of formulaicity: identifiability, fixedness, compositionality, and meaning shifts. To put it simply, we suggest that it is important to be aware that:

- language input contains not only individual words but also longer sequences, and these sequences are not random;
- (quite often) these sequences have to be used in a certain inflexible form, even if similar and grammatically correct forms are available;
- (quite often) their meaning cannot be derived directly from putting together all the single words that compose them;
- (quite often) single words in these sequences could have a meaning that is different (to a certain degree) from their typical use.

Since the definition of metalinguistic awareness proposed here is not limited to explicit metalinguistic knowledge, it is not expected that language learners would be able to explicitly describe formulaic types and the features mentioned above, but rather that being aware of these aspects they might allocate their attention accordingly when handling formulaic items. This observed sensitivity to the important aspects of formulaicity that results from a certain degree of understanding of their nature is what we regard as metalinguistic awareness. The way learners' sensitivity was measured in the present study is described in more detail in the methodology section.

As previously seen, the studies that looked into L2 learners' noticing of formulaic sequences or sensitivity to other formulaic features measured these abilities at the group level (with the partial exception of Martinez & Murphy, 2011). However, possible individual differences in

these aspects would make an important contribution to our understanding of formulaic competence.

In the master thesis (Moskvina, 2017), that created a foundation for the present study, I examined the learning of pragmatic formulaic sequences from watching L2 videos. In line with Martinez and Murphy (2011), the analysis of the post-viewing interviews[7] revealed that participants differed in their approach to the target items. While some learners focused on individual words within an expression to construct its meaning ("*I know this word, so I know this expression"*), others recognized them as a unique class potentially warranting special treatment and were therefore more cautious in interpreting the meanings of their individual constituents ("*I know this word, but it is an expression, I think it might mean something different"*). This varied perception of the same items could potentially suggest individual differences in learners' metalinguistic awareness of formulaicity, and it is worth further exploring whether these differences might lead to varied learning outcomes.

To our best knowledge, only one study attempted to explicitly measure individual differences in learners' (metalinguistic[8]) awareness of formulaic sequences. In the study by Guz (2014), 112 advanced English learners (L1 Polish), enrolled in a language teacher training program at university, were presented with two consecutive tasks. First, participants had to translate a text from L2 English to L1 Polish, which ensured their deep engagement with its semantic and structural organization. After handing in the translations, they were asked to highlight in the same text "groups of two or more words which they felt should be treated as one, complete

[7] Unpublished data

[8] It is called by the author simply "awareness", however, it is very closely linked to our definition of metalinguistic awareness.

unit rather than as individual, independent words" (Guz, 2014, p.171). The answers were then scored against the formulaic analysis of the text carried out by the author with reliance on several dictionaries. The results confirm generally low awareness of the presence of formulaic sequences in the input, with an average of two-thirds of the items not being noticed by the learners (12 out of 30 identified). Besides, the individual variation in the data was assessed. A relatively low variability was reported, however, participants were future language teachers in the second year of training, which suggests that they were used to deep linguistic analysis and metalinguistic terms, and as a result might have had a higher level of metalinguistic awareness than the general population, and less variation.

The qualitative analysis of participants' choices revealed that a number of features contributed to the perceptual saliency of the selected formulaic sequences: lexical and/or grammatical abnormality (appearance of low frequency words or atypical constructions, e.g., *go berserk*, or *lucky me*!), typographic salience (separation by commas or an exclamation mark, e.g., *worse luck!* or ,*suffice to say*,), and non-compositionality. The author associated non-compositionality with the lack of transparency, deeming non-compositional sequences semantically opaque for language learners. This account of compositionality seems to contradict Martinez and Murphy (2011) and Kim (2016), who found that language learners were not aided by this feature and consistently failed to recognize non-compositional idioms in the input. Several explanations could be put forward to reconcile these findings. First of all, unlike the previous two studies, Guz (2014) supposedly deals with known formulaic sequences (at least to a certain extent), as participants were able to consult unknown words with the instructor during the translation task and no confusion over formulaic meanings was reported. Thus, non-compositional forms could have become more salient if compared to the

corresponding meanings. Secondly, Martinez and Murphy (2011) as well as Kim (2016) explicitly focused on formulaic sequences consisting of high frequent, familiar words, which was not the case for the discussed study. While some items did contain only frequent potentially known words, e.g., *I am no angel, in my book, pick up*, others did not, e.g., *care a toss, keep your pecker up, frothing mad*. It could be probably suggested that it is the other features discussed by the authors, such as lexical and grammatical abnormality, that make the non-compositionality of these items more salient.

Fixedness of form (or structural rigidity) and phrasal verb-like appearance were also expected to predict the selection of formulaic sequences in the study, which was, however, not supported by the data.

Having discussed the concept of metalinguistic awareness, which deals with the ways learners might perceive essential properties related to meaning organization in formulaic language, we move to the statistical learning ability, which potentially reflects learners' sensitivity to the distributional properties of formulas.

1.2.3. Statistical learning ability

Previous studies have shown that language learners exhibit certain sensitivity to statistical features of formulaic sequences at the group level, with this sensitivity typically following a different pattern from that of native speakers (e.g., Ellis et al., 2008; Siyanova-Chanturia et al., 2011; Hernández et al., 2016; Supasiraprapa, 2019).

In fact, the usage-based approach to SLA goes as far as to claim that the very process of language learning is essentially the absorption of statistical patterns, proposing frequently repeated sequences as main language building blocks (Ellis, 1996). Therefore, further

investigation of possible individual differences in statistical learning and their effect on formulaic acquisition is an important direction for SLA research.

Statistical learning describes learning mechanism based on the detection and assimilation of the regularities present in the input (Bogaerts et al., 2022). Although it was initially proposed to explain the fast and stable rate of L1 acquisition in children, and the majority of studies in the field are still focusing on L1 acquisition, statistical learning ability has been shown to correlate with learning of some L2 aspects as well, e.g., reading skills (Brice et al., 2022; defined as literacy in Frost et al., 2013;), morphosyntactic patterns (McDonough & Trofimovich, 2016), general proficiency level (Ettlinger et al., 2016).

Unlike metalinguistic awareness, which involves conscious, deliberate attention allocation to various language elements and their features in the input, statistical learning is a purely implicit process that takes place during input processing without learners even realizing it.

A number of studies discussed individual differences in statistical learning ability (e.g., Siegelman & Frost, 2015; Siegelman et al., 2020; Isbilen et al., 2022), with the latest directions in this line of research emphasizing the importance of a domain specific view of this ability (Bogaerts et al., 2022).

To our best knowledge, only two studies so far have attempted to explore the potential connection between individual differences in statistical learning ability and processing of formulaic sequences. Isbilen et al. (2022) examined individual differences in L1 English speakers' sensitivity to statistically defined multiword chunks in artificial and natural (English) language. Two measures of statistical learning were applied: an artificial grammar learning task (AGL) and statistically-induced chunking recall task (SICR). Before the tasks

were administered, participants were exposed to an 11-minute-long stream of an artificial language, with the stream containing statistical patterns that were later used for testing. In the AGL task, they were presented with pairs of syllable strings and had to choose which string within a pair was a "word" from the artificial language they heard. In the SICR task, participants listened to six-syllable strings that were comprised of two "words" put together without a pause in between and had to write them down. Statistical learning measures were compared with participant processing of three-word English lexical bundles (e.g., *good to know* or *is really nice*). The processing task involved listening to a series of concatenations of four lexical bundles (e.g., *have to eat good to know don't like them is really nice*) and writing them down as accurately as possible. The results showed that participants' ability to recall lexical bundles correlated with their performance on the SICR task but not with their scores from the AGL task. Since the SICR task measured participants' sensitivity to statistical properties of the input, it was further concluded that acquisition of multiword combinations could be potentially explained in terms of statistical learning.

Kerz and Wiechmann (2019) investigated the role of statistical learning in L2. The study measured statistical learning ability of 60 advanced English learners through a number of statistical learning tasks and compared the scores with participants' performance at online processing of multiword sequences. The statistical learning measurements included 3 artificial grammar learning tasks (AVA: Auditory-Verbal-Adjacent, AVN: Auditory-Verbal-Nonadjacent, VNA: Visual-Nonverbal-Adjacent; respectively dealing with adjacent or non-adjacent dependencies in verbal or non-verbal stimuli presented in aural or visual form) and a probabilistic Serial Reaction Time task (SRT). Online processing of formulaic sequences was evaluated through a version of a grammaticality judgement task, where learners had to

determine (as quickly as possible) if the item on the screen is an acceptable expression in English or not. Target items were four-word sequences extracted from the spoken and academic (written) components of the Corpus of Contemporary American English (COCA). Two of the statistical learning tasks, AVA and SRT, were found to be significant predictors for the variation in processing of multiword sequences. Learners' scores at the AVA task were positively related to their reaction times for the sequences from the spoken register and negatively related to the processing scores for the academic formulaic sequences, with precisely the opposite relationship emerging for the SRT task. Although no explicit explanation of the directional differences is put forward, it could be inferred that they might be accounted for by the discrepancies in the presentation mode between the real-life exposure to the target items and that of the statistical learning measures, i.e., the AVA task measures learners sensitivity to the distribution patterns in aural stimuli, while academic formulaic sequences belong to the written component of the COCA and are therefore most likely to be encountered in written form. The opposite holds true for the SRT task. In general, the results reinforce the connection between the sensitivity to distributional patterns and language processing as well as offer some evidence of this connection being extended to the formulaic items. However, it is duly emphasized that statistical learning is of a complex multi-component nature and more studies are needed to further these claims.

While it has to be acknowledged that the aforementioned studies focus mainly on processing of formulaic sequences, the observed patterns could be expected to extend to learning. As Bishop (2004) pointed out, language learners might struggle to identify boundaries of formulaic sequences in the input, and it can be theorized that statistical learning ability might

facilitate this process as well as general form-meaning mapping, which in case of formulaic items would involve relating a certain word sequence to a certain meaning.

It is further suggested that metalinguistic awareness and statistical learning ability could be symbolically reflecting the claims of the two main approaches in formulaic language research, the usage-based and phraseological ones. Do learners need to be aware of different formulaic types and their features or could they just absorb them from the input based on their statistical properties? Comparing the individual contribution of metalinguistic awareness and statistical learning ability to learning of formulaic sequences we might get some insight into this.

We thus conclude the introduction of the language focus of the present dissertation and the learner-related factors that will be considered here due to their potential relevance for acquisition of the target language items. The following sections of the literature review are devoted to the selected learning context and important context-related factors.

1.3. Context-related factors

1.3.1. Introducing the context: audio-visual input in SLA

Input has been long recognized as a crucial factor in second language learning. However, in a traditional instructed context input is quite often limited to the few hours that learners spend in class or engaged with the homework, and even during these hours input quality might vary considerably (e.g., Northbrook & Conklin, 2018). Input scarcity can have an especially detrimental effect on learning of formulaic sequences, since convention, or "typical" language use is an important defining part of this language aspect, and not having sufficient

authentic exposure to target language might prevent learners from forming accurate intuitions about what this convention consists of.

The situation has changed drastically when with the recent widespread availability of TV shows and films in the original version through various streaming services language learners gained unlimited access to authentic and engaging input (Neuman & Koskinen, 1992; Koolstra & Beentjies, 1999; Lin, 2014).

Besides being an abundant source of rich input, L2 videos provide special conditions for language learning, as they combine two modalities of information presentation: audio and image. They are therefore traditionally referred to as multimodal or audio-visual input. The advantage of multimodality is often explained through the multimedia principle, which states that people learn better from the combination of words and pictures than from words alone (Mayer, 2009). The principle was put forward as a part of Cognitive Theory of Multimedia Learning. Although it was initially proposed for general learning in L1, most of its constituent principles have been shown to equally apply for SLA[9]. The theory draws on Paivio's Dual Coding Theory (Paivio, 1990) that argues the existence of separate channels for processing verbal and imagery information in the human mind, resulting in separate but referentially connected storages. Thus, if the same item is presented in the input in both verbal and imagery forms, two mental representations of the item are created instead of one, reinforcing its subsequent recall.

[9] Except for the redundancy principle that was reconsidered by the author to account for L2 learning as suggested by empirical data.

However, as Mayer emphasizes, a mere combination of visual and verbal components cannot ensure learning and efficient multimedia instruction needs to be designed carefully in order to promote learning.

This reflection is further supported by Sweller's Cognitive Load Theory (1994), which highlights the necessity of considering in instructional design the cognitive architecture of the human brain and its main components: long-term memory and working memory. According to him, learning occurs in certain stages: exposure to novel information, retrieval of relevant items, and their storage in working memory (WM); potentially followed by integration of those items into long-term memory. Since working memory, unlike its long-term counterpart, has limited capacity (Miller, 1956), Sweller suggests that the principal aim of instruction should be facilitating optimal use of learners' working memory without its overload.

In the case of audio-visual input this can be achieved with the addition of a third modality - on-screen text, or subtitles. Mayer et al., (2020) advanced the subtitle principle asserting that learners can rely on the on-screen text when audio processing is challenging or incomplete. As the text is available for processing for a longer time than the audio and can be revisited several times, it releases the strain on learners' working memory. In the same line of thought, Vanderplank (2016) claims that the three sources of information (audio, image and text) complement each other thus promoting efficient employment of learners' working memory resources. This is corroborated by the finding from Pattemore & Muñoz, (2020), where differences in working memory affected grammar learning in the groups watching L2 TV series without on-screen text, but not for those who were provided with L2 subtitles, and

those from Gass et al. (2019) where at the second part of the experiment language learners with higher working memory relied less on captions than learners with lower WM scores.

On-screen text can be presented in learners' native or target language (coinciding with the audio). The terminology varies across the studies, with the on-screen text in the native language called subtitles, L1 subtitles or interlingual subtitles; and that in the target language referred to as captions, L2 subtitles, or intralingual subtitles. In this dissertation (L1) subtitles and (L2) captions will be used correspondingly.

1.3.2. Language learning from audio-visual input

The agenda for studies on audio-visual input includes two main questions: what can be learnt through this type of input and what are the optimal conditions for taking the full advantage of its richness.

It has been demonstrated that exposure to audio-visual materials can promote learning of various language aspects, such as vocabulary (e.g., Pujadas & Muñoz, 2019; Fievez et al., 2020, Gesa & Miralpeix, 2022), comprehension (e.g., Fievez et al., 2020; Pujadas & Muñoz, 2020), grammar (e.g., Cintrón-Valentín et al., 2019; Pattemore & Muñoz, 2020), pragmatics (e.g., Khazdouzian et al., 2021; Barón & Celaya, 2022), and pronunciation (e.g., Wisniewska & Mora, 2020).

In an attempt to identify the optimal parameters for using L2 videos as multimedia instruction, researchers explored the factors that might enhance or hinder learning of these aspects from audio-visual input, considering learner related as well as video related variables.

The investigated video related variables included on-screen text presence (e.g., Sydorenko, 2010; Montero Perez et al., 2014), language of on-screen text (L1 vs L2, e.g., Peters et al.,

2016; Fievez et al., 2020), type of on-screen text (e.g., glossed captions: Fievez et al., 2023; bilingual captions: Wang & Pellicer-Sanchez, 2022), instructed vs incidental learning (e.g., Pujadas & Muñoz, 2019), time between viewings (e.g., Muñoz et al., 2022), input enhancement (e.g., Lee, & Révész, 2018; Pattemore & Muñoz, 2022).

The following learner related factors have been considered: age of learners (e.g., Muñoz, 2017), their proficiency level (e.g., Muñoz, 2017; Gesa & Miralpeix, 2022), individual differences in working memory (Matielo et al., 2018; Pattemore & Muñoz, 2020), language aptitude (e.g., Pattemore et al., 2021), attention (Suárez et al., 2021).

The present dissertation does not strive to add to the discussion of the on-screen text advantages and rather aims to build upon the existing evidence in order to explore the factors affecting learning further.

In order to establish a clear objective for our study, learning is operationalized as incidental learning, a process not involving any intent on the part of the learner. Therefore, pre-teaching, explicit instruction, or any other attention drawing activity that would deliberately shift learners' focus from meaning to form are not reviewed here.

1.3.3. Learning of formulaic language through audio-visual input

Previous research has shown quite consistently that audio-visual input, and captioned videos in particular, can be beneficial for incidental vocabulary learning. However, most of the vocabulary studies involved individual words and longer sequences unified under the term "formulaic language" have been brought into focus only recently.

Lin (2014) explored the validity of internet television for acquisition of formulaic language, comparing the distribution of formulaic sequences in the 7.68 million-word iTV corpus of

11 genres against the spoken component of the British National Corpus (BNC). Her findings suggest that the distribution patterns in the corpus accurately reflect the real-life use of formulas, as it contains the most frequent formulaic items and their frequency is directly proportional to that of the spoken BNC. The author therefore argues that internet television could assist language learners in forming valid intuitions about relative frequencies of formulaic items. It is of much importance for the full command of formulaic language, as a number of researchers pointed out that disproportionate use of formulaic sequences (overuse of certain items and underuse of others) contributes greatly to non-native like performance (e.g., Nesselhauf, 2005; Paquot & Granger, 2012).

Lin (2014) notes as well that the distribution patterns varied across the sub-corpora of different genres, with some genres representing the use of formulaic sequences in everyday speech more precisely than the others. Genre differences and their potential effect on learning are discussed in detail in the corresponding section below.

The learning potential of TV programmes identified in Lin's (2014) corpus study has been experimentally explored in several studies. For example, Frumuselu et al. (2015) investigated learning of some formulaic items from watching TV series "Friends" under a captioned or subtitled condition. Two groups of Spanish/Catalan speakers (n=18 for the subtitle group; n=22 for the caption one) were exposed to 13 episodes of the show over 7 weeks. Results show that participants learnt better from L2 captions than L1 subtitles. It has to be mentioned that though her description of the target items included idioms and phrasal verbs, no clear definition of formulaic language was provided, and the focus of the study was described through quite general terms of "informal expressions" and "colloquial language". Thus,

while it is clear that the study adds to our understanding of learning of formulaic sequences through audio-visual input, its exact contribution is difficult to assess.

Moskvina (2017) looked at noticing and acquisition of pragmatic formulaic sequences from exposure to the captioned videos of two genres, as well as the effect of input enhancement. Her findings suggested that input enhancement in the form of highlighting can promote learning of formulaic items and to some extent impact their noticing (as measured by eye-tracking), while genre effect was only present for noticing.

Puimège and Peters (2019a) researched learning of single words and formulaic sequences through uncaptioned audio-visual input. Twenty L1 Dutch English learners watched 30 minutes of a British reality TV show focused on business. The list of 20 target formulaic items was comprised of lexical collocations (9), phrasal verbs (4), idioms (3), grammatical collocations (2), a binomial (1), and a compound[10] (1), with all but three items occurring once in the program. The authors found that incidental learning of formulaic sequences from L2 viewing without on-screen text can occur at the level of form, but not meaning recall. It is duly mentioned that though the findings suggest potential positive impact of audio-visual exposure on the learning of formulas, it could have been mediated by the test effect. The other significant predictors of learning included vocabulary size (positive effect), mutual information (negative effect), item length in syllables (negative effect), and collocate-node relationship, with the highest gains associated with adjective-noun combinations and the lowest with verb-particle ones.

[10] The target item was described as such in the list, but its classification was not further commented on in the text.

The authors further explored the potential of audio-visual input for learning of formulaic language in Puimège and Peters (2020). Forty-two upper-intermediate Dutch learners of English watched a one-hour documentary without any form of the on-screen text. The learning of 56 formulaic sequences was measured at the levels of form and meaning recall. The number of their occurrence varied, with 8 items appearing more than once. The target items belonged to the same set of categories as in Puimège and Peters (2019a), except for one simile. The results show significant learning gains at both form and meaning recall, although the higher gains for form rather than meaning recall suggest the test effect, which was confirmed in the questionnaire administered at the end of the experiment. The pre-viewing meaning recall scores emerged as the strongest predictor for the form recall at the post-test. It is therefore inferred that language learners are able to successfully build upon their previous knowledge of the items when provided with meaningful comprehensible input. Confirming their previous findings, collocate-node relationship and prior vocabulary knowledge impacted learning outcomes as well. Item-related variables, such mutual information and transparency of the items affected learning at the level of meaning but not form recall. No significant effect of L1 similarity was found.

Majuddin et al. (2021) examined the role of captions, textual input enhancement, and repetition on the learning of multiword expressions[11] from watching an episode of the American sitcom "Fresh off the Boat". One-hundred and twenty-two Malaysian L2 English learners were assigned to one of the six experimental conditions, that varied in the following

[11] Although the authors draw an explicit distinction between the terms "multiword expressions" and "formulaic sequences", it is not consistently supported throughout the literature. Therefore, both terms are used interchangeably in this dissertation, including the description of the aforementioned study.

characteristics: captions/no captions, watching the episode once/twice, having the captions enhanced/unenhanced. The eighteen target items included mostly idioms and phrasal verbs, although they were not described as such in the study, since the authors defined multiword expressions in terms of frequency and mutual information. Each of the items appeared once in the episode.

The results support the beneficial effect of captions on the form recall of formulaic items at the immediate and delayed post-tests. Repetition, on the other hand, positively impacted the recall only immediately after the viewing, with no differences in its effect for no-caption, caption, and enhanced caption groups. In line with the previous studies on vocabulary learning from audio-visual input (Pujadas & Muñoz, 2019; Puimège & Peters, 2019a; Montero Perez, 2020; Puimège and Peters, 2020), vocabulary knowledge affected participants' performance at the immediate and delayed post-tests, with learners with bigger vocabulary size tending to score higher. Prior knowledge of the target items also led to higher scores at the immediate post-test. The enhanced caption condition significantly improved learning of the target items over no-caption, but not over the caption one, with the effect fading at the delayed post-test.

The role of textual enhancement in learning of formulaic sequences from audio-visual input was considered by Puimège et al. (2021) as well. The study analyzed processing (measured by eye-tracking) and intake (gains on form recall) of 22 formulaic items from watching 30 minutes of a captioned documentary. Participants were 28 Dutch learners of English, their profile as well as the target items characteristics were similar to those from Puimège and Peters (2019a; 2020). It was found that input enhancement leads to deeper processing of FS in the captions, which in turn, results in better learning. This effect, however, was mediated

by individual differences in the attention paid to the captions. The influence of item related variables (mutual information, frequency, subjective decomposability, etc.) was present for some of the processing measures.

The most recent study by Vu et al. (2023) compared incidental learning of collocations from two input modes: captioned videos and reading-while-listening. The general findings indicate that both modes can promote collocation learning. The authors believe that the absence of differences between the conditions suggests insignificant contribution of image to collocation uptake, with the opposite being true for individual words in the previous studies (e.g. Rodgers, 2018; Peters, 2019). L1 similarity was a significant predictor of learning gains for both input modes. Unlike prior studies, vocabulary size had a negative effect on learning from L2 videos, with bigger vocabulary size leading to lower gains. A possible explanation is offered suggesting that more proficient learners might focus more on following the plot rather than paying attention to individual language items. The effect was positive for the reading-while-listening mode.

To summarize the findings obtained so far, it has been demonstrated that audio-visual input is potentially a rich authentic source of formulaic language and exposure to L2 videos can promote its learning, especially under captioned conditions. Several formulaic types have been brought into focus, with the research covering mainly collocations, phrasal verbs, and idioms. Learner-related factors have been discussed as well, however, the discussion was limited to individual differences in vocabulary size and prior knowledge of the items.

The expansion of the research agenda to include other formulaic types consistently as well as a broader spectrum of individual differences might improve our understanding of acquisition of formulaic sequences from the context in question.

Another aspect that has been considered in previous research on acquisition of formulaic sequences through audio-visual input is the type of context. This was explored through the comparison between captioned and no captioned videos, and the effect of input enhancement. The context-related factor of genre, that remains underresearched in the field of audio-visual input is further discussed below.

1.3.4. Genre in audio-visual input

While a wide range of learner and video related factors have been covered in an attempt to determine the optimal design of multimedia instruction, it is surprising that the genre of audio-visual materials was not consistently brought to light. Genre is a default property of any L2 video, however, its use by researchers is mostly limited to identifying the materials selected for the study, predicting their possible suitability for learners (e.g., choosing an "entertaining" genre of situational comedy in Rodgers, 2013) or serving as a covariate partially explaining the results (Garza, 1991; Neuman & Koskinen, 1992). Only a few studies discussed below considered genre as an individual context related variable that might affect learning. The lack of research into genres is confirmed by Vanderplank (2016), who shares his personal pedagogical experience with various genres and highlights the implication of genre selection on the learning process. For example, he reports having difficulty in using soap operas and dramas in general for language teaching. When presented with a soap opera his students failed to take the viewing activity seriously despite its learning potential having been emphasized by the teacher. The author further infers that a certain degree of engagement with the plot and characters is needed for following dramas. News programs are regarded by Vanderplank as one of the most optimal genres, since they focus on the current affairs and learners can easily relate to the content as well as to follow it provided that the exposure is

consistent. Documentary and sitcom are described as engaging as well, with a concern being voiced over the visual appeal of documentary as potentially distracting from the language aspect. Although of more an observational nature and aimed specifically at using L2 videos as a guided classroom activity, these findings add to our conception of genres and their potential role in learning.

1.3.5. Identifying the learning potential of genres

Vocabulary demands of different genres were investigated in two corpus studies by Webb and Rodgers (2009a; 2009b). In their first study the researchers analyzed the scripts of 318 movies that were either British or American, and belonged to one of the 11 genres: action, animation, comedy, suspense/crime, drama, horror, romance, science fiction, war, western, and classic, in order to identify the vocabulary size necessary to reach 95% and 98% of coverage, i.e. the vocabulary size that would account for the 95% and 98% of the word types in the scripts. It is suggested that these percentages might determine lower and upper boundaries for adequate comprehension of movies, at reasonable and ideal levels correspondingly, and therefore are useful for estimating how easy or challenging the analyzed movies might be for language learners in terms of comprehension. The results indicate that the lexical demands of different genres were quite comparable at the level of 95% coverage, with knowledge of 3000 most frequent words plus marginal words and proper nouns being sufficient to reach this level for 8 out of 11 genres (slightly higher coverage was provided for western, classic, horror, and drama). Knowledge of 4000 most frequent words plus marginal words and proper nouns was necessary to get 95% of coverage for the genres of war, animation, and action. More significant differences emerged at the coverage level of 98%,

with vocabulary size needed to reach this level varying from 5000 words[12] (horror, drama, crime) and 6000 (western, romance, classic) to 7000 (comedy, action, science fiction) and 9000 (war), with the highest of 10000 (animated). It can thus be inferred that genres differ in their vocabulary demands for language learners. Horror, drama, and crime appeared to be the least demanding genres in the study, and animation was shown as the most challenging one. On-screen text is suggested as a tool for lowering vocabulary demands. Certain differences between American and British movies were observed at 98% of coverage, with British movies being slightly less challenging, which might be explained by the usage of the vocabulary lists based on the British National Corpus.

Although the relationship between coverage and comprehension is not completely straightforward and certain variation within a genre might be present, coverage can be interpreted as an indication of a potential difficulty level of audio-visual materials. Overall, since comprehension plays an important role in incidental learning (e.g., Zeeland & Schmitt, 2012; Montero Perez et al., 2014), we can conclude that these findings can be discussed in terms of learning potential of different genres, i.e., their potential to promote incidental learning. The authors also point out that the importance of following the dialogue closely might vary according to genre, for instance, missing a few key phrases in the crime genre might result in more confusion about the plot development than missing an obscure joke in the comedy.

Vocabulary load of different TV genres was further examined in Webb and Rodgers (2009b). The study involved 88 American and British TV programs, belonging to one of the 6 genres:

[12] All the provided coverage numbers include marginal words and proper nouns. The phrase is not repeated for the sake of brevity.

news, drama, situational comedy (sitcom), older programs, children's programs, and science fiction. The results indicate that for British programs drama, sitcom, and science fiction genres were less demanding than the news one, and there was considerable difference in lexical demands between the episodes of the news genre, which might be explained by the shortness of individual episodes and variation in the topics they cover. For American genres, children's programs were the least challenging, followed by older programs, sitcom, and drama. News and science fiction appeared as the most challenging genres. Vocabulary demands varied between the episodes as well, leading authors to conclude that along with general genre characteristics, those of concrete episodes should be taken into account as well in the choice of multimedia materials. Learning potential of genres was further discussed in terms of low frequent words and their re-occurrence. The percentages of reoccurring words from the most frequent 5000- 14000 word families were compared across the genres. Since the vocabulary size of approximately 3000 is assumed as sufficient to follow the TV programs, the authors suggest that the potential for incidental learning of new vocabulary lies within the aforementioned frequency range and is reinforced by the number of times the supposedly new items reappear. In this respect the most beneficial for learning were children's programs, with the lowest percentage of low-frequent words appearing only once, and the highest of re-occurring at least 5 times. The other genres that were shown to be favorable for incidental vocabulary learning included news, science fiction, and older programs. However, it is admitted that the 4000 most frequent words level might hold more realistic potential for language learners, though the proportional distribution of the words from this frequency band across the genres is not further commented on.

The two studies provide a great insight into vocabulary demands of different genres that might be helpful in defining their learning potential. It has to be noted though that the classification of the genres is somewhat arbitrary. For instance, "I love Lucy" and "Gilligan's Island", listed as older programs for being filmed in the 60s-70s, are also commonly recognized as sitcoms. The TV series "C.S.I." and "24" classified as drama in Webb and Rodgers (2009b) follow the work of investigators and easily fall in the category of crime genre that was identified as an independent genre in Webb and Rodgers (2009a) and analyzed separately from the drama one. Although it is quite common for TV programs to be a blend of several genres, a stricter categorization of genres might be needed for any strong conclusions about their learning potential.

Another approach to evaluating learning potential of different genres was proposed by Rodgers (2018). The study suggests that since image is expected to provide additional support in learning of new words through audio-visual input, any differences in this aspect might affect learning outcomes. Two types of TV programs were compared: narrative (a fantasy adventure drama) and documentary, as they were suspected to differ considerably in the imagery patterns. The results confirm the author's hypothesis, showing that though imagery in both genres effectively cooccurred with the audio creating conditions for potential learning of concrete nouns, these conditions were more favorable in the documentary genre. While in the documentary 65.2% of the imagery occurred simultaneously with the aural form of the target words, and 72.3% within the 10 second timeframe, for the narrative program these measures were at 15.6% and 29.3% correspondingly. The reoccurrence of the target items differed as well between the two genres, with an average of 11.3 times for documentary against 7.6 for the narrative TV. It is thus inferred that specific characteristics of the

documentary genre, such as heavy reliance on imagery and in-depth exploration of one topic with extensive vocabulary repetition might create beneficial conditions for learning, although other variables, e.g., learners' interest in the said topic, need to be taken into account as well.

Finally, learning potential of different genres for the formulaic part of vocabulary was analyzed by Lin (2014). The study suggests that although internet television in general can be regarded as a valuable source of formulaic sequences accurately representing their real-life distribution, the degree of accuracy varies across the genres. Factual, drama, and comedy genres were reported to exhibit the most similar patterns to the everyday use of formulaic language, while music, learning, and religion were found to be the most dissimilar. It is further developed that genre should be considered in designing multimedia instruction, along with other important parameters of audio-visual input, such as on-screen text, its language, learners' characteristics, etc.

In conclusion, the analysis of audio-visual materials from different genres indicates that they vary in a number of aspects and this variation potentially results in distinct learning conditions provided by each genre.

It is interesting to note that only two video aspects have been reviewed in connection with learning potential of different genres: image and vocabulary demands. However, genre is a complex phenomenon that also includes soundtrack, content organization and plot development within and between the episodes, presence and type of action (competing with the dialogue or not), the direction of the dialogue (other characters or the audience, as in the news programs), etc. Different combinations of these characteristics might impose different demands on learners' attentional resources, leading to different learning outcomes. Thus, for

the purposes of this dissertation we define **genre** as *a set of unique properties that create specific conditions for learning.*

1.3.6. Experimental testing of learning potential

Despite the ongoing discussion on the learning potential across the genres, its experimental testing has been extremely limited.

Gilabert et al. (2018) investigated noticing (measured by eye-tracking) and initial vocabulary learning from the exposure to short clips of 4 genres: documentary, edutainment (TED talks), police procedural (crime), and sitcom. In a counter-balanced, within subject design learning of 40 items (10 per genre, 9 individual words and one formulaic sequence) was measured at the levels of meaning recognition (pre- and post-tests) and form recognition (post-test only) for 41 Catalan-Spanish learners of English. The highest number of word forms and meanings was recognized from the documentary, followed by sitcom, crime, and TED talks. Genre also had an effect on noticing, with more attention paid to the target items in sitcom and documentary. Besides, learners skipped less words in TED talks and documentary than in crime and sitcom. A weak correlation between noticing and learning was reported as well. The results tentatively suggest documentary and sitcom as the most beneficial genres for initial vocabulary learning, though participants also paid consistent attention to the subtitles in TED talks.

Suarez et al. (2021) further explored the same dataset to examine the role of individual differences in working memory, attention, inhibition control, and vocabulary size on the initial learning of the target items. Vocabulary size was identified as a significant predictor of learning, however its effect varied across the genres, with learners relying less on their

previous vocabulary knowledge while watching the documentary. It corroborates Rodgers' (2018) claim that documentary provides stronger imagery support than other genres. No significant impact of the cognitive variables on the initial learning was found.

Learning of formulaic sequences under different genres was touched upon in Moskvina (2017). Noticing and intake of pragmatic formulas were compared between two genre groups, sitcom and TED talks. Genre did not seem to have a significant effect on learning, it did however affect noticing (measured by eye-tracking). Language learners engaged more with the target items in the captions when watching TED talks than they did for sitcom, which could be explained by the absence of any distracting action in the edutainment genre.

As per the literature available the evidence obtained so far provides certain experimental support for the distinct learning potential of different genres and establishes direct links between the discussed learning potential and actual learning. However, the diverse learning conditions defined by the intrinsic properties of different genres and their effect on learning outcomes have not been sufficiently explored so far. More experimental studies are needed to draw any definitive conclusions and the present study aims at filling this gap.

Additionally, contextual differences associated with different genres might have a more pronounced effect on learning of formulaic sequences than on learning of individual words, as they would require deeper contextualization, especially in case of pragmatic formulaic sequences. This will be further explored in this study, which we describe in the following section.

1.4. Present study

In the present study we aim to further our understanding of the acquisition of formulaic language and the factors that might affect such process.

Since learning of formulaic sequences involves acquiring not only the immediate language form but also the convention associated with it, consistent exposure to authentic materials in L2 is highly beneficial for the acquisition of this language aspect. Learners need to be in contact with the target language to be able to form intuitions about what are the "common" ways to express certain meanings. Audio-visual input, easily accessible through various streaming platforms, has been shown as a valid method to provide this contact.

Previous research demonstrated that exposure to audio-visual input can result in incidental learning of a number of language aspects, with very promising results emerging from vocabulary studies. Several recent studies have proved that this beneficial effect holds for the formulaic part of the vocabulary, indicating that L2 subtitled (captioned) videos might be the most efficient way to achieve it. It has to be mentioned, however, that the overall number of the studies remains very low, and more research is therefore needed to explore the potential of audio-visual input for acquisition of formulaic language.

These studies brought into focus a number of different formulaic sequences, yet the range of discussed types is limited and somewhat inconsistent. Two main approaches could be observed in the choice of target items: analyzing all the variety of formulaic types that naturally occur in the selected video materials or focusing on a specific pre-selected category. In both cases the formulaic types of idioms, collocations, and phrasal verbs seem to be favored, with the other items making very limited appearance (binomials) or not being included at all (lexical bundles, pragmatic formulaic sequences). Since previous research

showed that processing of various formulaic categories might differ, and it is not quite feasible to find a balanced distribution of different categories within an episode in order to carry out a between-types comparison, it was decided here to focus on one specific formulaic class.

For this purpose, pragmatic formulaic sequences have been chosen as the main linguistic focus of the present research. To our best knowledge, no studies so far investigated learning of pragmatic formulas in the audio-visual context. This research gap is quite surprising, as these formulaic sequences are highly dependent on the context and therefore could benefit the most from the exposure to various communicative situations containing their authentic use.

However, as it has been noted before, providing learners with rich authentic context is not always enough. Although the specific features of audio-visual input have proved to make it an effective tool for promoting incidental learning, researchers are still working to determine the most advantageous ways to employ it. Since the primary objective of authentic TV programs is to entertain, inform, or produce a work of art, they are not designed as learning materials and not necessarily follow the criteria for efficient multimedia instruction. They merely provide massive amount of input that, especially in instructed learning contexts, would not otherwise be available.

Certain aspects of L2 videos that might affect learning of formulaic sequences have been analyzed so far, including presence of captions, their language, and input enhancement. Based on previous findings, the present study focuses on captioned videos as a more learning beneficial form of audio-visual input. At the same time, such a basic property of any video material as genre remains virtually unexplored both in formulaic studies as well as in general

vocabulary ones. While corpus studies suggest that different genres might have different learning potential, the experimental testing of this claim has been extremely limited. It is further proposed here that the differences in learning potential between genres are not confined to vocabulary of the scripts. Each genre includes a set of audio, video, and narrational parameters that are combined in very particular ways, and these combinations define specific learning conditions associated with the genre. These conditions might vary in the demands they put on viewers' attentional resources and the support they provide for various aspects of language. Therefore, the quality of multimedia instruction involving different genres might vary as well. Differences in learning conditions associated with the context are of much importance for pragmatic formulaic sequences, as these formulaic items need deeper contextualization and are more likely to get ignored in case of attention overload than other language items.

It has to be further acknowledged that even with the most efficient multimedia instruction, learning outcome still largely depends on the learner. So far only two learner-related factors were considered in acquisition of formulaic sequences from audio-visual input: vocabulary size and prior knowledge of the target items. The exploration of other individual differences might provide deeper insight into the learning process in this context. Since the analysis of previous research on formulaic acquisition suggests that language learners often fail to treat formulaic items as a special category, we propose metalinguistic awareness and statistical learning ability as the learner-related variables that potentially reflect two main formulaic dimensions, statistical regularities and meaning organization peculiarities. The comparison of an implicit cognitive statistical learning ability with more explicit metalinguistic

competence can provide interesting insights into the involvement of corresponding processes in incidental learning from captioned videos.

Besides, having a closer look at learners' metalinguistic awareness across formulaic types and features might give us better understanding on what they consider to be formulaic and what aspects of formulaicity they struggle with.

To conclude, it can be claimed that learning of formulaic sequences through captioned L2 videos warrants further research in terms of the items considered as well as learner- and context-related factors.

To fill in the identified research gap and to deepen our understanding of acquisition of formulaic sequences, the following research questions were put forward:

Research question 1. Is the learning of formulaic sequences affected by the type of conditions provided by different audio-visual genres (at the example of sitcom, TED talks, drama, and action genres)?

Since previous research into this issue is scarce, we cannot propose any clear directional hypothesis for this research question. However, we believe that the selected genres would provide different learning conditions, which, in turn, might result in higher or lower learning outcomes.

Learning is operationalized here as intake, based on the interactionist model of SLA suggested by Gass (1997). The model defines intake as comprehended input stored for the further processing, however, since it is difficult to establish whether certain elements in the input were comprehended or not, we opt for a broader definition proposed by Reinders:

> "Intake is a subset of the detected input (comprehended or not), held in short term memory, from which connections with long term memory are potentially created or strengthened." (Reinders, 2012, p.28).

The rationale for operationalizing learning as intake lies in the potentially high relevance of initial stages of learning in the context of audio-visual input, especially in case of incidental learning. We cannot expect learners to exhibit productive knowledge of formulaic sequences after seeing them on screen a few times, we can, however, hope that they would extract some useful information about these sequences for further processing and assimilation. In addition, focus on the initial stages of the learning process might demonstrate a more pronounced effect of metalinguistic awareness, with learners who are more aware of formulaic language being able to extract the needed information more efficiently. More details on the research design, including the test chosen to measure intake are provided in the methodology chapter.

Research question 2. Do individual differences in statistical learning ability mediate learning of formulaic sequences from audio-visual input?

Based on the limited evidence from the studies that examined the connection between statistical learning ability and processing of formulaic sequences, we expect learners with a higher level of the said ability to benefit from the suggested audio-visual treatment more, since it would potentially be easier for them to segment the input and correctly identify the boundaries of formulaic sequences. In turn, that might help them to create stronger form-meaning connections.

Research question 3. Does metalinguistic awareness of formulaicity mediate learning of formulaic sequences from audio-visual input?

To this end, a multicomponent test of metalinguistic awareness will be designed. We expect a faciliatory effect of metalinguistic awareness on the intake of formulaic sequences from audio-visual input, however, since this question has not been raised before, this part of the present study is mostly exploratory. Metalinguistic awareness of formulaicity, defined here as sensitivity to the essential (meaning-related) properties of formulaic language might aid learners in selecting relevant information about the use of formulaic items from the input.

Research question 4. Are there any interactions between learner-related and context-related factors?

It is possible that learning conditions associated with certain genres might require more assistance from learners' statistical learning ability and metalinguistic awareness than others. Additionally, the effects of such learner-related variables as age, vocabulary size, and prior knowledge of the target items should be explored for each genre in order to ensure that they do not vary depending on the genre (all of them, except age, were addressed in previous research on formulaic acquisition as well).

Research question 5. Does L2 learners' metalinguistic awareness of formulaicity vary across formulaic types and features?

Finally, we hope that the data collected with a new measure of metalinguistic awareness of formulaicity will give us insights into learners' idea of what comprises formulaic language, and which formulaic features they are more aware of.

Thus we complete the review of the literature relevant for the objectives of this doctoral dissertation. The implementation of the introduced research agenda is further discussed in the next chapter.

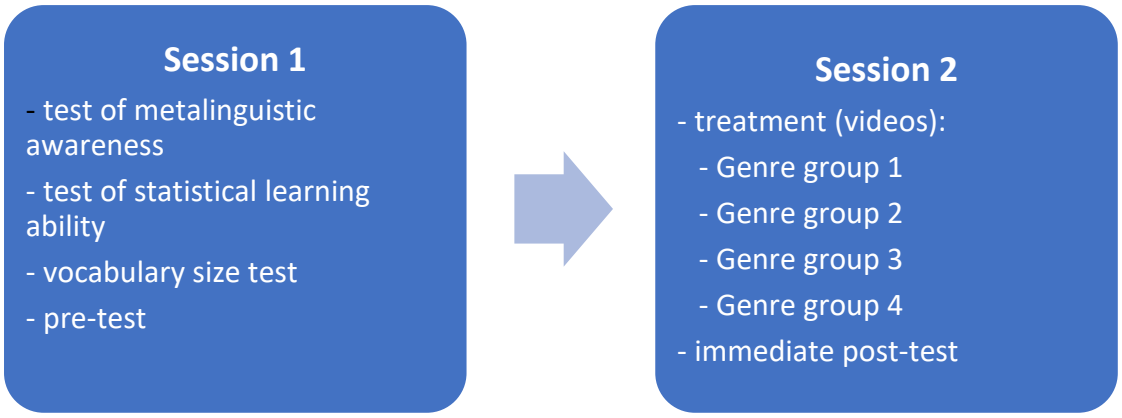
Chapter 2. Methodology

This chapter is devoted to the methodology of the present dissertation. It explains in detail the research design of our study, instruments and materials applied to answer the research questions and the procedure that was followed, as well as the participants from whom the data was collected. The chapter is composed of 10 sections, with each of the sections focused on a certain element of the research design. It starts with the overall description of the design, followed by the detailed account of all its constituents and ends with the preliminary analysis section.

2.1. Experimental design

The study followed a pre-test/treatment/post-test design and was divided into 2 sessions (Figure 1). For the treatment session, participants were assigned to one of the 4 experimental groups. No control group without any on-screen text was included as previous studies convincingly demonstrated the beneficial effect of captions, and the purpose of our research was to further explore video properties and their potential effect on learning rather than to contribute to the captions vs. no captions debate. The nature of the recruitment process and the difficulty of collecting data after the COVID-19 pandemic did not allow for a no-treatment, pre- post-test control group.

Figure 1. Experimental design



2.2. Participants

Participants of the study were L1 Catalan/Spanish learners of English recruited from 8 Escoles Oficials d'Idiomes (or EOIs, official language schools in Catalonia) and from the Faculty of Education of the University of Barcelona (Early Childhood Education degree). The recruitment targeted learners at A2-B2 proficiency levels as they were considered optimal for the planned treatment to be beneficial. Subjects' actual level was then measured more precisely during the study. As metalinguistic awareness is one of the key variables of the study, it was ensured that none of the participants had any linguistic training/degree or a language related job (language teacher, translator, etc.), since this could have resulted in the enhanced level of metalinguistic awareness. Information about the study was distributed through the teachers, and learners could sign up directly in class or via an online Google form. To avoid revealing the research objectives, the study was advertised with the very generic goal of "practicing English by watching videos". All participants were provided with feedback on their results and a certificate of participation, those from the University of

Barcelona received 10% of the course credit as a reward. The appropriate ethical procedures were followed in all cases.

A total of 377 English learners signed up for the study and were provided with personalized materials, 114 of them finished all the tasks and underwent the treatment (31 males, 83 females). Three participants were not able to perform the statistical learning ability task due to unsolvable technical issues, but they were not excluded from the study as they completed the rest. All 114 participants of the final sample were native speakers of Spanish, 107 of them reported to be bilingual in Spanish and Catalan, 7 reported only speaking Spanish. From Catalan/Spanish speakers, 5 participants stated to be trilingual with one of the following languages as their third mother tongue: Chinese (1), Galician (1), Bulgarian (1), Arabic (1), Farsi (1). The mean age of the sample was 31.85 (range 17-73), with the mean vocabulary size 4750.41 (range 2009-8078).

High heterogeneity of the sample is partially caused by the decision to work with the real-life classroom demographics rather than focusing exclusively on university students. A typical class at an EOI includes learners of various ages and backgrounds, and even though the classes are organized by a proficiency level, learners' actual language performance lacks uniformity. A careful grouping procedure was implemented to ensure that the differences in age and proficiency did not interfere with the results (see a detailed description in the section *Procedure* below).

2.3. Target items

As mentioned in the previous chapter, the language focus of this dissertation is formulaic sequences. The following 6 formulaic sequences were chosen as the target items for the present study:

- Let's face it
- Good for you![13]
- Give me a break
- Don't get me wrong
- Never mind
- What's the big deal?

The target items belonged to pragmatic formulaic sequences, also known as speech formulas, pragmatic routines, or situation bound utterances (SBUs). The choice of this formulaic type is explained by its high reliance on the context, which is expected to be helpful in contrasting contextual support provided by different genres. Besides, unlike other formulaic types, they usually form complete phrases or sentences, thus ruling out the necessity of controlling for the differences in the immediate lexical and grammatical context. The criteria for the selection of the particular sequences were the following:

1. Consisting of highly-frequent words that would be familiar to the participants.

[13] Although this item might seem less challenging than others, it did not appear to be the case in Moskvina (2017), where it was included as a distractor, therefore, it was further explored here as a target item.

2. Not entirely compositional, i.e. not having their meaning directly derived from their constituents, as in, for example, in SBUs *"do you have a minute"* or *"nice to meet you"*.
3. Potentially unknown (or at least not completely acquired) at the proficiency level of the target population (A2-B2). This was determined by three English teachers with linguistic training.
4. Occurring naturally and in its most typical use in the selected genres.

First, a list of potential items was composed, based on the formulaic sequences that appeared in previous studies (e.g., Bardovi-Harlig, 2009; Moskvina, 2017; Kecskes et al., 2018) and three fellow teachers' expertise. It was then narrowed down to the items that could be found in the TV shows of the four experimental genres and most fully met the criteria from above.

The number of target items was limited to six in order to ensure the optimal length of the viewing experience (approximately 30 minutes) and control over the conditions under which they are presented to the participants.

2.4. Proficiency test

V_YesNo v1.0 was administered to measure participants' proficiency in English. It is a vocabulary size test based on the Eurocentres Vocabulary Size Test (Meara & Jones, 1990).

The test contained 200 words that were presented on the screen one at a time. Participants were asked to press the "yes" button if they knew the meaning of the word, and "next" if they did not. It was emphasized that in case of any uncertainty about the answer, the option "next" should be selected. One hundred of the presented items were real English words, while the

other one hundred were invented ones. The real words were a stratified sample from the vocabulary of 10.000 words and the invented ones were included to correct for guessing.

Participants received the test in the form of a hyperlink that took them to the lognostics website (Meara, 2022). Their answers were scored automatically by the algorithm and possible scores ranged from 0 to 10000. Learners were asked to write down the final score or take a screenshot of it. The test manual was used for the interpretation of the scores.

Despite its main focus on vocabulary knowledge, the YesNo test has been shown to be a valid measure for overall proficiency, with its scores fairly correlating with other language skills (Meara & Miralpeix, 2016). The decision to use a vocabulary size test to estimate participants' proficiency was also justified by the fact that vocabulary is the primary focus of the study. Its quick administration and simple structure were considered beneficial as well, as they helped to reduce the cognitive demand of the pre-viewing tests and were optimal for online data collection.

2.5. Knowledge of the target items

In order to estimate the learning of the target items, pre- and post-tests were designed based on the Vocabulary Knowledge Scale (VKS; Wesche & Paribakht, 1996). The original scale was modified for the SBUs following Bardovi-Harlig (2008). The pre- and post-tests were identical and included 16 pragmatic formulaic sequences: 6 target ones and 10 distractors. Five distractors were of the same difficulty level as the target items and the other five were considered easier and already familiar for potential participants. The easy distractors were included to motivate participants and give them a sense of achievement, while the more

challenging ones were supposed to deflect participants' attention from the target items. None of the distractors appeared in the videos during the treatment session.

The test was administered in Google Forms, and learners received a corresponding hyperlink with the rest of the materials. They were asked to evaluate their knowledge of each of the sequences based on the following 5-point scale (English version):

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don't know what it means.

☐ I have seen this expression before, and I think it means:

Please, write a synonym or translation/explanation

_____

☐ I know this expression. It means:

Please, write a synonym or translation/explanation

_____

☐ I can use this expression in a context:

(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

_____

_____

For the full version of the test, see Appendix A.

Participants could choose a Spanish or Catalan version of the test with the scale translated in their preferred language.

As VKS was designed to probe into initial stages of vocabulary learning, it seemed well suited for the purpose of the present study. Another advantage of this test is that it allows to

measure slight changes in participants' knowledge, which is favorable for experiments involving limited exposure.

The original scoring system from Wesche and Paribakht (1996) was applied. The scores ranged from 1 (the expression is completely unfamiliar) to 5 (the expression can be used in a context semantically and pragmatically appropriately), with other scores being 2 (the expression is familiar but the meaning is unknown), 3 (the meaning is known but without certainty) and 4 (the meaning is known). Incorrect translations were assigned a score of 2, incorrect use in a context was scored as 4 if the correct translation was provided and 2 if the translation was not correct either.

The answers were scored by the researcher, in case of any ambiguity in translation or use, a final ruling was made by a native speaker with linguistic training (English or Spanish/Catalan depending on the case).

2.6. Audio-visual materials

2.6.1. Genres

Four genres were chosen for the purposes of the present study: **situational comedy** (sitcom), **edutainment** (TED talks), **drama** (teen drama), and **action** (superhero action).

The selection was carried out through the following steps:

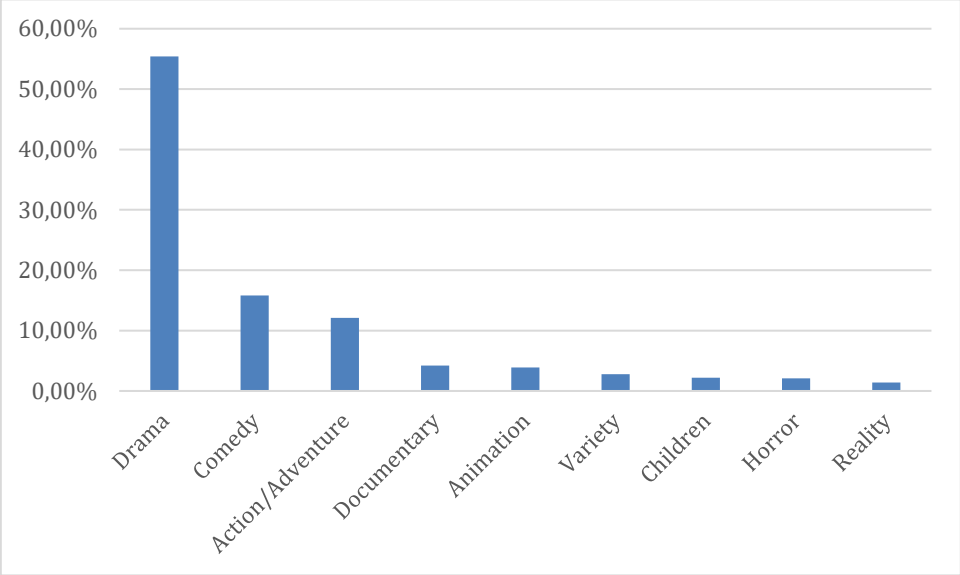
1. Analysis of the genres used in previous research.
2. Analysis of the viewing demands in Spain.
3. Analysis of the potential candidates to identify the most suitable for the purposes of the study.

The following major genres have been consistently appearing in the literature on audiovisual input:

- Documentary (e.g., Garza, 1991 *(combined with news genre)*; Winke et al., 2010; Gilabert et al., 2018; Peters & Webb, 2018; Peters, 2019)
- Comedy/sitcom (e.g., Vanderplank, 1988; Garza, 1991; Webb & Rodgers, 2009a; Webb & Rodgers, 2009b; Lin, 2014; Gilabert et al., 2018; Majuddin et al., 2021)
- Animation (e.g., Garza, 1991; Webb & Rodgers, 2009a)
- Action (e.g., Webb & Rodgers, 2009a; Webb, 2011 *(as a subcategory of drama genre)*)
- Drama (e.g., Garza, 1991; Webb & Rodgers, 2009a; Webb & Rodgers, 2009b; Lin, 2014; Rodgers & Webb, 2020)
- Music (e.g., Garza, 1991; Lin, 2014)
- Children's (e.g., Webb & Rodgers, 2009b; Lin, 2014)
- Science fiction (e.g., Webb & Rodgers, 2009a; Webb & Rodgers, 2009b)
- News (e.g., Webb & Rodgers, 2009b; Lin, 2014)
- Edutainment/TED talks (Moskvina, 2017; Nurmukhamedov, 2017; Gilabert et al., 2018)

To identify the most popular genres in Spain, the Global Television Demand Report (Parrot Analytics, 2021) was consulted (Figure 2).

Figure 2. Digital series demand by genre, Q2 2021 (Spain)



The choice of genres for the study was affected by some practical and ethical issues as well. For example, despite the fact that children and older programs appeared to be the most favorable genres in terms of vocabulary load and learning potential in Webb and Rodgers (2009b), these genres could not be used here since they had a very specific target audience (children programs) or were not enough "mainstream" (older programs). Similarly, we could not use the horror genre, which had relatively low vocabulary demands in Webb and Rodgers (2009a), since it could have negatively impacted participants' physical and mental health.

Likewise, documentary appeared in a number of previous studies and ranked as the 4th most popular genre in Spain in 2021, however, it was not included in the final selection. This genre was not considered to be a prolific context for natural appearance of pragmatic formulaic sequences due to the lack of variety in the communication situations this genre traditionally depicts.

On the other hand, TED talks (or edutainment) were not among the genres preferred by Spanish viewers but were regarded as an important genre due to its unique set of characteristics and its frequent use in teaching-related research (e.g., Nguyen & Boers, 2019; Madarbakus-Ring, 2020).

Beside popularity and previous mention in research, the selected genres were considered sufficiently distinct, i.e., they varied across a number of aspects and therefore created clearly different learning contexts. For the detailed analysis of the genre properties, see the section *Videos* (2.6.3.) below.

Overall, the selection of the genres was supported by the previous research and the industry data.

2.6.2. Selected TV programs

The following TV programs were used to operationalize the selected genres:

Sitcom: "Joey" and "Friends".

Edutainment: various TED talks.

Drama: "Gossip Girl", "90210".

Action: "The Flash", "Legends of tomorrow", "Arrow", "Daredevil" (the movie).

(For the full list of all the episodes appearing in the study see Appendix B)

These TV shows were considered representative for the chosen genres as they exhibited most of the common features associated with each genre and did not have a double classification. For example, sitcoms like "Scrubs" and "Parks and Recreation" are also classified as medical drama and mockumentary correspondingly. Combination of several genres within a TV

program can result in a specific set of features that will not be shared with other shows belonging to one single genre. Thus “Scrubs” has a lot of medical terms which is unusual for a sitcom, and “Parks and Recreation” includes many scenes where the characters address the audience directly, imitating the documentary genre.

The initial selection was limited to the TV shows that would certainly be unfamiliar for potential participants (e.g. shows “Joey” and “90210” that were not broadcast on central Spanish TV and are not quite well-known in general). However, the choice was restricted by the natural occurrence of the target items and the fact that the most typical representatives of each genre tend to be the most popular ones (it is especially the case for superhero action). Thus, the final selection included the shows that might be familiar to participants with familiarity addressed in the analysis.

Combining different shows was not regarded as a problem as there is considerable overlap or similarity between the shows within a genre. In the case of the action genre, all of the shows but one (“Daredevil”) belong to the same fictional universe and frequently have cross-over episodes or references. “Joey” is a spin-off of “Friends”, while “Gossip girl” and “90210” have considerable similarities in the narrative.

2.6.3. Videos

The audiovisual treatment involved 12 videos for each of the 4 genres (48 video clips in total). Each clip was approximately 2 minutes (120 seconds), with the length balanced across the genres as shown in Table 2 below.

Table 2. The length of videos across the genres (in seconds)

Genre	Mean (SD)	Min	Max	Total length
Sitcom	118.92 (11.2)	102	136	1427
TED talks	116.25 (9)	101	133	1395
Drama	119.92 (10.8)	103	135	1439
Action	118.67 (11.3)	101	134	1424

The length was considered optimal for sufficient contextualization of the target items without overextending the overall watching time. In order to facilitate comprehension, the clips were cut in a way that ensured they included meaningful scenes. Each video contained only one target formulaic sequence, hence there were 2 clips per each of the items. Target items appeared approximately in the middle of the clip, where we assumed participants' concentration was at its highest level. The mean difference between the exact middle of the clip and the time of target item appearance was 7.7 seconds for sitcom videos (SD=4.6), 7.3 (SD=3.7) for TED talks, 7 for drama (SD=4.2), and 7.6 (SD=4.8) for action. All differences in length and target item appearance were matched across the genre, i.e. if in one of the genres a target item appeared in a shorter clip 10 seconds after the middle point, the corresponding videos of the other genres had similar characteristics.

The selected videos were representative of the genre they belonged to. For instance, all the superhero action clips included an action scene of some intensity (two of them contained a suspense build-up to a fight which is another common feature of the genre). Despite some TED talks incorporating elements of interview, dance, or animation, all the chosen videos

had the most typical format of delivery: one speaker addressing the audience, no action and no auxiliary materials involved (except for one video that contained a couple of slides). All the drama videos included an emotional dialogue, while all sitcom clips focused on a humorous situation and had a laugh track.

All the videos had original English audio and English captions. The TED talk videos were downloaded from the official website with default captions embedded. For the other genres, captions were created in Aegisub Advanced Subtitle Editor 3.2.2. using Netflix Timed Text Guidelines for style and timing (Netflix, n.d.) and embedded into the clips using VLC media player (VideoLan, 2006).

To estimate vocabulary demands of the genres and to ensure the accessibility of the video materials for participants, lexical coverage of the clips was measured with RANGE software (Heatley, et al., 2002) and British National Corpus/Corpus of Contemporary American English word lists (Nation, 2017). Table 3 demonstrates the coverage (%) provided by the knowledge of 1000, 2000, 3000 words for each video of each genre, with and without proper nouns and marginal words.

Table 3. Coverage provided by the knowledge of the most frequent 3000 words for each video in the study

N	Sitcom		TED talks		Drama		Action	
	Without proper nouns and marginal words %	With proper nouns and marginal words %	Without proper nouns and marginal words %	With proper nouns and marginal words %	Without proper nouns and marginal words %	With proper nouns and marginal words %	Without proper nouns and marginal words %	With proper nouns and marginal words %
V1	1. 94.06[14] 2. 96.04 3. 96.04	1. 97.69 2. 99.67 3. 99.67	1. 84.02 2. 90.21 3. 93.3	1. 87.63 2. 93.82 3. 96.91	1. 91.9 2. 92.71 3. 93.52	1. 95.54 2. 96.35 3. 97.16	1. 94.92 2. 96.61 3. 96.61	1. 97.17 2. 98.86 3. 98.86
V2	1. 92.31 2. 94.06 3. 95.11	1. 96.86 2. 98.61 3. 99.66	1. 86.50 2. 95.09 3. 97.24	1. 86.50 2. 95.09 3. 97.24	1. 90.67 2. 93.78 3. 95.11	1. 95.12 2. 98.23 3. 99.56	1. 83.05 2. 88.98 3. 91.1	1. 85.59 2. 91.52 3. 93.64
V3	1. 88.21 2. 91.06 3. 91.06	1. 93.5 2. 96.35 3. 96.35	1. 85.63 2. 89.15 3. 91.5	1. 86.21 2. 89.73 3. 92.08	1. 89.63 2. 93.31 3. 93.64	1. 94.64 2. 98.32 3. 98.65	1. 89.50 2. 94.07 3. 94.53	1. 92.24 2. 96.81 3. 97.27
V4	1. 94.41 2. 95.11 3. 95.46	1. 95.81 2. 96.51 3. 96.86	1. 84.23 2. 93.24 3. 97.74	1. 84.23 2. 93.24 3. 97.74	1. 89.76 2. 92.47 3. 93.07	1. 92.47 2. 95.18 3. 95.78	1. 89.30 2. 92.98 3. 94.65	1. 91.98 2. 95.66 3. 97.33
V5	1. 90.74 2. 93.46 3. 94	1. 92.37 2. 95.07 3. 95.63	1. 85.45 2. 90.54 3. 92.36	1. 87.26 2. 92.35 3. 94.17	1. 88.02 2. 93.15 3. 94.62	1. 89.98 2. 95.11 3. 96.58	1. 89.80 2. 92.94 3. 94.51	1. 91.36 2. 94.5 3. 96.07
V6	1. 90.92 2. 94.39 3. 95.65	1. 93.38 2. 97.55 3. 98.81	1. 88.72 2. 91.01 3. 93.37	1. 91.83 2. 94.12 3. 96.48	1. 91.62 2. 94.22 3. 95.09	1. 95.09 2. 97.69 3. 98.56	1. 87.77 2. 89.9 3. 93.62	1. 89.37 2. 91.5 3. 95.22
V7	1. 84.41 2. 86.78 3. 87.46	1. 93.23 2. 95.6 3. 96.28	1. 79.63 2. 87.04 3. 91.98	1. 82.1 2. 89.51 3. 94.45	1. 90.62 2. 93.85 3. 95.02	1. 93.84 2. 97.07 3. 98.24	1. 95.16 2. 96.77 3. 96.77	1. 95.16 2. 96.77 3. 96.77

[14] Numbers 1, 2, 3 stand for 1000, 2000, and 3000 most frequent words.

V8	1. 90.51 2. 93.67 3. 93.67	1. 93.28 2. 96.44 3. 96.44	1. 89.31 2. 92.2 3. 94.51	1. 90.47 2. 93.36 3. 95.67	1. 92.61 2. 93.75 3. 93.75	1. 97.44 2. 98.58 3. 98.58	1. 83.26 2. 86.34 3. 86.78	1. 85.9 2. 88.98 3. 89.42
V9	1. 89.21 2. 94.19 3. 94.6	1. 93.77 2. 98.75 3. 99.16	1. 82.13 2. 88.4 3. 94.04	1. 84.32 2. 90.59 3. 96.23	1. 88.28 2. 91.58 3. 92. 68	1. 93.04 2. 96.34 3. 97.44	1. 91.50 2. 93.85 3. 94.14	1. 93.56 2. 95.91 3. 96.2
V10	1. 87.79 2. 91.95 3. 92.99	1. 92.21 2. 96. 37 3. 97.41	1. 85.19 2. 94.76 3. 97.54	1. 85.5 2. 95.07 3. 97.85	1. 89.86 2. 94.06 3. 94.41	1. 93.71 2. 97.91 3. 98.26	1. 89.66 2. 93.68 3. 94.83	1. 92.25 2. 96.27 3. 97.42
V11	1. 87.87 2. 89.18 3. 91.15	1. 95.08 2. 96.39 3. 98.36	1. 89.43 2. 94.87 3. 96.08	1. 89.43 2. 94.87 3. 96.08	1. 89.45 2. 93.12 3. 95.41	1. 92.66 2. 96.33 3. 98.62	1. 85.47 2. 90.2 3. 91.89	1. 89.53 2. 94.26 3. 95.95
V12	1. 89.6 2. 91.2 3. 92	1. 95.6 2. 97.2 3. 98	1. 78.4 2. 83.73 3. 88.76	1. 82.84 2. 88.17 3. 93.2	1. 94.22 2. 95.24 3. 95.58	1. 97.62 2. 98.64 3. 98.98	1. 83.73 2. 89.47 3. 89.47	1. 90.91 2. 96.65 3. 96.65

As it can be seen from the table, most videos reached at least 95% coverage necessary for adequate comprehension (Webb & Rodgers, 2009a) at the level of 3000 word families plus proper nouns and marginal words. The exceptions are 4 TED talks videos (V3, V5, V7, V12) and 2 videos belonging to the superhero action genre (V2 and V8). The detailed analysis of these videos, however, showed that none of the words not covered by the 3000 of the most frequent word families were essential for understanding the content of the clips, and many of them would be easily recognizable for participants because of their cognateness with Spanish or widespread use in popular culture, e.g., *glucose* (Spanish: *glucosa*), *diabetes* (Spanish: *diabetes*), *latina* (actual borrowing from Spanish), *girlfriend* (a compound but a highly common word) in TED talks; *assassin* (Spanish: *asesino*), arrogant (Spanish: *arrogante*), vibrate (Spanish: *vibrar*), velocity (Spanish: *velocidad*), *cookie* (common word) in superhero action. Besides, most studies on vocabulary coverage dealt with "raw" materials, i.e., the materials that did not include any additional vocabulary support for learners. In our

research[15], all the videos were prefaced with brief introductions that placed them into context, which significantly lightened their vocabulary demands (for more details about the introductions, see the section below). We can thus claim that the video materials were accessible for our participants, as all of them but 12 had a vocabulary size of 3000 words or higher, and it is assumed that the twelve learners who scored between 2000 and 3000 received sufficient support from the introductions (for more details about the introductions, see the next section).

The differences in coverage between the genres reflect the general intention in preparation of the video materials, which was to make them as accessible and as comparable as possible without eliminating the inherent genre differences. For example, the original subtitling style for TED talks is quite distinct from the rest, but that is the style learners will be exposed to if they watch this genre at home or in the classroom, so it did not seem reasonable to transform it.

These inherent genre differences are described more thoroughly in Table 4 below. It includes both general properties as well as those that might be of importance for acquisition of pragmatic formulaic sequences.

[15]It also involved captions, which is considered as a possible tool to lessen vocabulary demands in Webb and Rodgers (2009a).

Table 4. Description of genre properties

Property	Sitcom	TED talks	Drama	Action
Total number of words	3534	3595	3569	2919
Words needed to reach 95% coverage[16]	2000 (96.83%)	3000 (95.62%)	2000 (97.1%)	2000 (95.2%)
Words needed to reach 98% coverage	4000 (98.1%)	6000 (98.02%)	4000 (98.24%)	5000 (98.45%)
Density of speech (words/min)	148.6	154.6	148.8	123
Lexical complexity (%)	1.67	2.87	1.21	2.24
Action type	Complementary action (dancing, fighting, exercising, etc.)	No action	Complementary action (dancing, fighting, exercising, etc.)	Action filled poses (chasing, fights), action competing with the dialogue
Plot organization	Each episode has a separate story, but there is an overlap.	Each episode has a separate story, no overlap.	Each episode contributes to the main story, considerable overlap.	Each episode contributes to the main story, considerable overlap.
Type of narration	(Mostly) dialogue	monologue	(Mostly) dialogue	(Mostly) dialogue
Required background knowledge	Cultural references, humor	Some familiarity with the	Cultural references,	Cultural references, familiarity with

[16] Both coverage % include marginal words and proper nouns. Due to the small samples, the coverage % without these categories are considerably skewed and therefore not informative.

		topic/context (varies)	familiarity with the characters	the characters and the fictional universe
Contextual clues	Gestures and facial expressions, intonation, interlocutor's reaction, clearly identified communicative situation	Gestures and facial expressions, intonation	Gestures and facial expressions, intonation, interlocutor's reaction, clearly identified communicative situation	Gestures and facial expressions, intonation, interlocutor's reaction clearly identified communicative situation

Lexical complexity was operationalized here as percentage of low-frequent words, i.e., 5000 frequency band and higher. Density of speech was calculated by the division of the total running time by the total number of running words. Finally, contextual clues describe the information that can be used by learners to interpret the meaning of pragmatic formulaic sequences.

2.6.4. Video sequence

Since our study did not involve intact episodes of the selected TV programs, and rather focused on short individual clips, these clips were combined in a video sequence along with some auxiliary elements in order to create a smooth viewing experience for the learners. The structure of the sequence is shown in Figure 3. The video sequence started with a slide containing a brief explanation of the structure of the viewing experience, the TV shows and/or the genre involved. Participants were also instructed to enter the full screen mode and not to press anything in order to move forward as the sequence progressed automatically. The explanation remained on the screen for 30 seconds until it was replaced with a "Let's

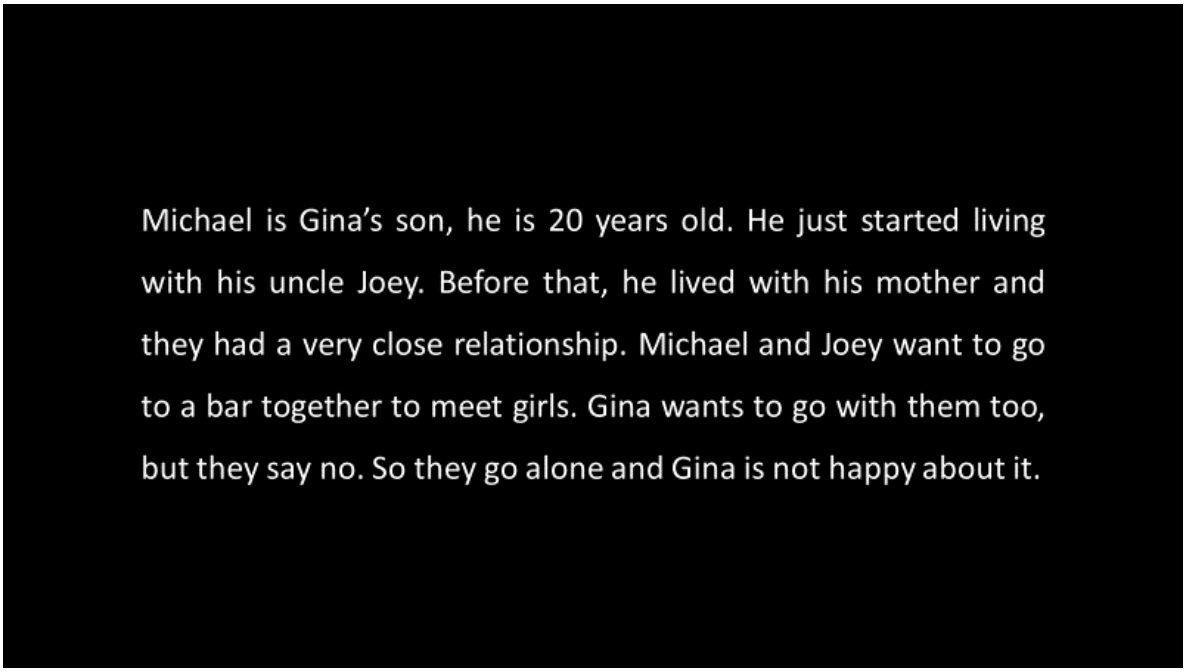
start!" slide that signaled the beginning of the viewing. There was one video sequence per genre, which contained all the 12 clips of that genre plus the described auxiliary elements.

As the clips within the sequence were essentially unrelated, there was a short introduction slide before each clip that explained the situation and the characters that appeared in the scene (for sitcom, drama, and action genres). For the TED talks, the speaker and the topic of the talk were identified. All introduction slides had a black background and white font, with the text organized in 5 lines. The introductions were delivered in English, and it was ensured that the vocabulary and grammatical structures of the introductions were accessible for target proficiency levels. If a video contained any language material that was considered challenging for the learners but important for understanding (cultural references, complex proper nouns, specialized vocabulary), it was mentioned, translated, or explained in the introduction. Each introduction slide stayed on for 25 seconds and then faded away as the video started. An example of an introduction slide could be seen in Figure 4.

After each video clip two questions appeared automatically, one about the content and one about the familiarity of the video (both in English). Comprehension questions were included to ensure participants' focus on the meaning and had two answer options: "yes" and "no", with the choice made by means of a mouse click. The responses were scored automatically, and subjects received immediate feedback with correct answers marked green and incorrect ones in red. In the familiarity question learners were asked to type "yes" or "no" depending on whether they had seen the video before. The questions were not timed and the introduction for the next video appeared after participants answered both of them.

Figure 3. The structure of the video sequence

Figure 4. Example of an introduction slide



Total length of the sequence was 29 min 27 sec for sitcom, 28 min 55 sec for TED talks, 29 min 39 sec for drama, and 29 min 24 sec for action. Within a sequence, the two clips containing the same target item immediately followed each other to maximize the limited exposure. The order of the clips was not randomized due to the technical limitations and was the same for all the participants of the same genre group.

The instruction and introduction slides were created in Microsoft PowerPoint and combined with the video clips in OpenShot Video Editor. The final sequence was then uploaded to Edpuzzle, a free educational resource for creating video lessons, where the questions and the

correct answers were added. The experimental links used in the study can be found in Appendix C. Edpuzzle has a number of features useful for collecting data online. For example, learners cannot fast-forward or skip the videos, and if they leave the webpage the video stops until they are back. It is also necessary to answer the questions to progress through the sequence. After (or during) the completion, the researcher can check how much time participants spent on a sequence and whether they watched any clip more than once.

The video clips were piloted with 3 L1 Catalan/Spanish learners of English with the proficiency level within the target ones (B1). The introductions and the instructions were piloted separately with two learners of the similar profile to ensure accessibility of the language and timing. The entire sequence was piloted with two native Catalan/Spanish speakers: one highly proficient and trained in linguistics, and one with the similar characteristics to potential participants.

During the experimental session, participants received an Edpuzzle link with the video sequence of the assigned genre. A personal Edpuzzle account was created for each participant to facilitate the access to the website and guarantee their identifiability. Learners could log in using the provided credentials or their Google account.

2.7.Metalinguistic awareness of formulaicity

To measure participants' metalinguistic awareness of formulaic language, a multi-component test was developed. It was composed of 5 tasks that focused on the 4 essential formulaic features established in the literature review: identifiability (2 tasks), meaning shifts (1 task), lexical fixedness (1 task), and compositionality (1 task). Cronbach's alpha of 0.6 showed an acceptable level of internal consistency, confirming that all 5 components tap into the same

construct of metalinguistic awareness. However, its position at the lowest boundary of the accepted range supports our intuition about discernable differences in their nature.

As we wanted learners to draw mainly from their metalinguistic awareness rather than purely linguistic knowledge, the accessibility of the language for the target population was ensured with two native English speakers trained in teaching and applied linguistics. In the cases where the use of potentially unknown language could not be avoided, "*I don't know*" option was introduced to prevent guessing.

The design of the test is described in details below. The full version of the test can be found in Appendix D (with the instructions translated in English for convenience).

2.7.1. Formulaic items

The test included all main types of formulaic sequences: idioms, collocations, lexical bundles, binomials, phrasal verbs, and pragmatic formulas (or speech routines). The presence of the items varied across the sub-tasks to avoid ambiguous cases, as explained below in the corresponding sections.

Overall, a special effort was made to avoid L1 similarity, that is, to ensure that the target items did not have any direct formulaic equivalents in Spanish or Catalan. This was tested with two native Catalan/Spanish speakers, and only one item (Task 1) was reported to have a reversed binomial equivalent (*day and night* vs *noche y día*). Its possible effect was addressed during the analysis, and since its exclusion or inclusion did not alter the results, it was kept as a target item.

Most of the target formulaic items were chosen from previous studies involving corresponding formulaic types or dictionaries. Collocations used in the test were mostly

restricted ones unless any specific task required otherwise. Lexical bundles were selected using the frequency criteria suggested by Biber et al. (2004): they appeared in the COCA 40 times per million words. Binomials were reversible with a strong preference for one order (Mollin, 2012), as irreversible binomials often have idiomatic meaning that would make them a special category combining the features of binomials and idioms. A possible exception is "*lost and found*" (Task 2), which was included as potentially familiar to learners. Idioms and pragmatic formulaic sequences were chosen according to the specific task requirements.

The detailed description of each task can be found in the corresponding sections below.

2.7.2. Task 1. Identifiability (text)

The task was partially based on Guz (2014), and its objective was to evaluate participants' ability to identify formulaic sequences in the input.

For this purpose, learners were given a short text to read (327 words). The text was based on a reading exercise for A2 level suggested by British Council LearnEnglish (British Council, n.d.). The text was modified to ensure that it contained several examples of all formulaic types considered in the present study. The final version incorporated 3 clear examples for each of the following sequence types: lexical bundles, collocations, binomials, idioms, phrasal verbs, and pragmatic formulas (or speech routines). Participants were told to read the text and highlight groups of two or more words that they considered to be a fixed unit rather than individual words. It was emphasized that they should not be discouraged by non-familiarity with the term "*fixed unit*" and should give honest answers following their own interpretation. No examples were provided to avoid priming, since the exposure to formulaic items could have affected learners' selection, especially with highly homogeneous types

(e.g., phrasal verbs). The test was piloted with and without examples and there were no significant differences in performance between the two versions (for more information about piloting, see the corresponding section below).

2.7.3. Task 2. Identifiability (list)

Unlike the previous task that was aimed at testing learners' ability to identify formulaic sequences in a more naturalistic situation (reading a text), this one evaluated the same ability in a more controlled setting: a finite list of expressions.

Participants were presented with a list of 24 expressions: 12 formulaic sequences (binomials, idioms, collocations, lexical bundles, phrasal verbs, speech routines; 2 sequences each) and 12 free expressions modelled after their formulaic counterparts ("*banana and honey*" for binomials, "*have apples*" for collocations). Learners were asked to highlight the expressions that they considered "fixed" and only choose those they were sure about.

2.7.4. Task 3. Meaning shifts

The objective of the task was to evaluate participants' sensitivity to meaning shifts within collocations. It included 18 items, 9 target ones and 9 distractors. Each item consisted of 4 sentences, with each of the sentences having one word marked **in bold**. Participants had to match the two sentences with the similar meaning of the marked words. The distractor items contained free combinations or unrestricted collocation only (e.g., *sad film, leave a bag*). For the target items, one of the sentences contained a meaning-shift unit, i.e., a collocation where one of the words obtained a different meaning as a part of collocation compared to its common use as an individual word (e.g., *run a shop* vs *run fast*), with the word appearing **in bold** being the one affected by meaning shift. The other three sentences included:

- one sentence with a word **in bold** that had the same/similar meaning to the "new" meaning resulting from the in meaning shift (the correct answer), (e.g., *run a shop* vs *manage a project*);
- one sentence with a word **in bold** that had the same/similar meaning to the common individual use of the word affected by the meaning shift (a distractor) (e.g., *run a shop* vs *escape from prison*);
- one sentence with a word **in bold** that either
    - could potentially replace the word affected by the meaning shift in the provided context but did not have the same/similar meaning (a distractor) (e.g., *run a shop* vs *rob a shop*).
    - Or could appear in the contexts similar to those typical of the word affected by the meaning shift in its common individual use (a distractor) (e.g., *deep trouble* vs *dark sea*).

The strategies for creating the distractors were based on Gierl et al. (2017).

Since the immediate context played an important role in this task, the instructions emphasized its importance in making the choice. Participants were also instructed to select the "I don't know" option in case they did not understand some of the sentences.

A full example of an item from this task can be found below.

1.
    a. My brother **ran** a shop here when he was younger. ☐
    b. Two men **robbed** a supermarket last night. ☐
    c. Maria **managed** this project well last year. ☐
    d. Hector **escaped** from prison five years ago. ☐
    e. I don't know. ☐

2.7.5. Task 4. Fixedness

The objective of the task was to measure participants' sensitivity to fixedness of formulaic items, i.e., their restricted or altogether absent variability.

It included 12 pairs of sentences, with the sentences within a pair differing only in one word or word order (for binomials). For each pair participants were asked to choose the acceptable variant, in case they considered both variants acceptable they were instructed to mark their choice accordingly. If learners were unsure about the meaning of the sentences, they had to mark the *"I don't know"* option.

Eight of the 12 pairs contained a formulaic sequence of one of the following types: binomials (2 pairs), idioms (2 pairs), speech routines (2 pairs), and collocations (2 pairs). Lexical bundles and phrasal verbs were excluded as these types cannot be considered representative of this feature. Within each pair, one sentence included an original, non-modified sequence (e.g., *piece of cake* or *take it easy*) and in the other one a word in a sequence was substituted for a synonym (e.g., *slice of cake* or *take it simple*). The selected formulaic items had a high degree of lexical fixedness and did not allow for substitutions, thus only the original sentence was considered acceptable (which was confirmed by native English speakers during piloting). It has to be mentioned that for the binomials the changes were syntactic rather than lexical, i.e., the reversed word order (*here and now* vs *now and here*). Although it deals with syntactic rather than lexical fixedness, it was considered more representative of the formulaic character of this item type.

The other four pairs were free combinations modelled after the target FS types. For these items a substitution of a word with a synonym (e.g., *you haven't replied to my email* vs. *you*

haven't answered my email) was perfectly acceptable, so participants were expected to select both variants.

All the items appeared in a clear context that would help learners in case of doubt, e.g., even if they were not fully familiar with the idiom "*a piece of cake*", they could deduce its meaning from the sentence "*The exam was so easy, a piece of cake!"*.

An example of an item from this task can be found below.

1. I can't talk now, I'll call you back. ☐
   I can't talk now, I'll call you around. ☐
   I don't know ☐

2.7.6. Task 5. Compositionality

The objective of the task was to measure participants' awareness of the non-compositional nature of formulaic items.

To this end, learners were presented with sentences that contained **phrases in bold**. For each case, participants had to decide whether it was possible to understand the meaning of the whole phrase knowing the meanings of all the words in it. The answer was provided by checking a "yes" or "no" box. It was emphasized that learners were not asked whether they can understand the phrases at the moment of testing but should rather imagine seeing them for the first time. If the meaning of a phrase was unfamiliar, they were advised to check the translation provided below (in Spanish or Catalan depending on the version).

The task included 24 items: 3 collocations, 3 idioms, 3 phrasal verbs, 3 speech routines (pragmatic FS), and 12 distractors. Clear non-compositional cases were chosen for each of the type, with compositionality of collocations identified by comparing the mutual

information of the phrases obtained by replacing one of the words in the collocation with a similar word (Lin, 1999). Lexical bundles and binomials were excluded as they tend to be fully compositional. The distractors were modelled after the 4 formulaic types appearing in the task.

An example of an item from this task can be found below.

1. The party downstairs really **raised the roof** last night.
   ☐ Yes
   ☐ No
   Translation: [hizo mucho ruido]

2.7.7. Task order

The order of the individual tasks within the test was not randomized between participants since the two identifiability tasks had to appear at the very beginning to prevent priming. Thus, the order was the same and learners were instructed to follow it without any modification. For the same reason the identification task involving a closed list of expressions followed the one based on a text.

2.7.8. Piloting

Before deciding on the final version, the test of metalinguistic awareness went through several rounds of piloting. During the first round, the test was administered with 7 native speakers of English. Five of them were trained linguists and two had advanced degrees in other fields. Their feedback on the instructions, accessibility of the language and clarity of the selection options was then incorporated in the second version of the test. This version was piloted with 10 L1 Spanish/Catalan learners of English enrolled in a B1 course at two language academies (five of them were provided examples in Task 1. and five were not).

Learners were asked to indicate any unfamiliar or confusing language, as well as any doubt that they might have. According to their responses, a number of ambiguous items were changed and the structure of Tasks 3. and 4. was optimized. The last round of piloting involved 8 English learners with a similar profile. The results were considered satisfactory and therefore the final version of the test was established.

Participants of the main study received the test in the form of a Word file. They completed it in a preferred text editor (Microsoft Word, OpenOffice Writer or Google Docs) and sent it back to the researcher via email. Learners could choose between the Spanish and Catalan versions, with the instructions appearing in the corresponding language.

The test was scored in the following way: all correct answers were assigned the value of 1, with incorrect ones registered as 0. Some task-specific decisions were made, as described below.

For Task 1 (identifiability in the text), firstly the pre-selected target formulaic sequences were scored and only this data was used for the item analysis to keep the balanced number of items across formulaic types. However, as the task involved a text suitable for low-intermediate learners, it contained many frequent constructions and phrases. Some of them were either not considered to be clear examples of formulaic language or were not recognized by the approach adopted in this dissertation (but distinguished in some others, e.g., Erman and Warren's (2000) reducibles: *let's, I'm*). We thus felt that some of participants' choices, although not coinciding with our initial selection, needed to be acknowledged. Learners received an extra point for recognizing each unrestricted collocation (e.g., *great idea*), lexical bundle (e.g., *sounds like*) or at least 3 grammatical constructions or reducibles (to ensure consistent recognition of the class, e.g., *is going to*). No differences emerged between the

scores with and without extra points, so the former was used in the further analysis as a more detailed measure (for more detail, see preliminary analysis).

For Task 3 (meaning shifts), only the target items were scored. As the task was meant to measure learners' sensitivity to meaning shifts, we believed that analyzing free combinations would not necessarily probe into the same ability. Besides, all "*I don't know*" answers were scored as missing values. Since participants were explicitly instructed to choose this option in case they were not familiar with the expressions, we assumed that these answers showed the inadequacy of linguistic knowledge of certain items and not the lack of the metalinguistic awareness. Therefore, no firm claims about such responses could be made and they could not be treated as truly incorrect answers.

The "*I don't know*" answers in Task 4 (fixedness) were marked in the same way. In addition, if learners selected a modified formulaic item as the only acceptable variant (e.g., *did a mistake* instead of *made a mistake*), it was scored as a missing value for the same reason of insufficient linguistic knowledge of the item.

2.8.Statistical learning ability

An Auditory-Verbal-Adjacent (AVA) task was used to measure participants' statistical learning ability. It is an artificial grammar learning task that involves recognizing adjacent probabilities in a stream of verbal stimuli presented in an aural form. Previous research showed that higher performance on the AVA task might be associated with greater sensitivity to the statistics of multiword units in the spoken register (Kerz & Wiechmann, 2019).

The AVA task used in the present study was designed after Palmer and Mattys (2016). It included two parts: familiarization phase and testing one. At the beginning of the

familiarization phase participants were informed that they would listen to an artificial language for 6 minutes and had to try to figure out what the words of the language were. They were instructed not to take any notes and listen attentively; the following testing phase was mentioned as well. During the familiarization phase participants were exposed to a stream composed by the concatenation of four invented words: *pemadovi, lasokachu*, *tinugo*, and *rebufi*. The words appeared in a pseudorandom order that did not allow for immediate repetitions. Each word was repeated 108 times at a speech rate of 4.17 syllables per second, making the stream total length equal to 432 words. The transitional probability between the syllables within a word was 1.00, while the transitional probability across the word boundaries was 0.33. This means that the syllable "pe" could only be followed by the syllable "ma" to form the word *pemadovi,* while the syllable "vi" could be followed by either "la", "ti", or "re", depending on the word that appeared next in the stream. There were no pauses or acoustic boundary cues between the words in the stream, so the only way participants could segment it properly was by relying on the transitional probabilities, and namely by comparing those within and across the words. The three syllable words were included to ensure that the rhythmical structure of the stream could not be used as a segmentation cue.

The familiarization stream was immediately followed by the testing phase that included 27 trials, presented in an audio form one at a time. Each trial was a pair of words and participants had to specify which word in a pair belonged to the language they had just heard. The selection was made by pressing the left or right arrow key on the keyboard for the first and second words correspondingly. The interval between the words within a trial was 500 ms, and between trials 1000 ms. Participants had 5000 ms to answer in each trial, after that the next trial started automatically. The first three trials were practice ones and included non-

words from another stream that participants did not hear during the experiment. The other trials belonged to one of the 3 types: word vs. non-word (W-NW, 8 trials), word vs. part-word (W-PW, 8 trials), and part-word vs. non-word (PW-NW, 8 trials). The third type was included to equalize the number of times each string type appeared during the testing. Part-words were strings of syllables across the word boundaries, e.g., *bufilaso* is a part-word formed by the last syllable of the word *re**bufi*** and the first syllable of ***laso**kachu*. The non-words were strings of syllables that existed in the stream but never appeared in that particular order, e.g. the non-word *chuvibugo* is comprised of the syllables from the words *lasoka**chu***, *pemado**vi***, *re**bu**fi*, and *tinu**go*** that could not possibly sequence in that way according to the rules of the invented "language". Both part-words and non-words included 2 strings of two syllables and 2 three-syllable ones, thus two words within each test trial were always matched in syllable length. The correct answer placement in a trial (first vs. second word) was counterbalanced within participants. However, the order of the trials was not randomized between the subjects as we were not interested in comparing trial types and it was assumed that any peculiarity it contained would affect all the subjects equally.

The audio files for familiarization and testing phases were identical to Palmer and Mattys (2016, stream A), as they were kindly shared by the authors. They were synthesized by the text-to-speech MBROLA software (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) using an English male diphone database (for more detailed information, see Palmer & Mattys, 2016). The "accent" was considered appropriate for the present study as it focuses on learners of English.

The task was created using OpenSesame 3.3.3. (Mathôt, Schreij, & Theeuwes, 2012) and adopted for online application with OSWeb extension (version 1.4.). It was then uploaded to

MindProbe, a free JATOS server that hosts online experiments. MindProbe allows for creating of personalized accounts for participants, where researchers can monitor their progress and store the results. After piloting with three Spanish/Catalan speakers the task was modified according to the feedback (clarity of the instructions, timing, etc.).

During the experimental session, the task was shared with the learners in a form of a hyperlink that would take them directly to the MindProbe website. The task would start automatically after the subject's confirmation of readiness. Participants could choose a Spanish or Catalan version with the instructions presented in their preferred language. The responses were scored automatically as the correct answer for each trial was specified in the algorithm. Only word vs. non-word and word vs. part-word trials were used for the analysis, they are further on referred to as “statistical learning measure 1” and “statistical learning measure 2”, respectively. The non-word vs. part-word trials are comparing two types of strings that were not statistically defined as words of the artificial language during the familiarization phase, so it would be difficult to assign the correct answer within the pair.

2.9. Procedure

Due to the uncertainty caused by the COVID-19 restrictions, the data collection was carried out online. After signing up for the study, either online via a Google form or through their teacher, participants received an email with the materials for the first session. It explained the structure of the session and described the required equipment (computer/laptop and headphones) as well. For each task, general instructions and approximate time of completion were specified, with more detailed guidance provided within the tasks themselves. Participants were told to follow the indicated order of the tasks, respond honestly without

consulting any external resources, and were asked not to look for the correct answers after completing the session. They were also instructed to do all the parts of the session on the same day without any interruptions in the middle of a task. The following order of the tasks was established:

1. **Task 1**: test of metalinguistic awareness
2. **Task 2**: test of statistical learning ability
3. **Task 3**: vocabulary size test (v_YesNo)
4. **Task 4**: pre-test of the knowledge of the target items (VKS) + background information

The order was not randomized and kept the same for all the participants, as this version was considered optimal for avoiding any interference between the tasks and balancing the cognitive load of the session. Tasks 1 and 2 were considered the most challenging and thus were put at the beginning of the session to ensure participants' highest concentration capacity. Besides, to prevent any interference from the vocabulary appearing in the other tasks (especially in the pre-test consisting of formulaic expressions), the test of metalinguistic awareness was administered as the very first task. As v_YesNo test does not allow to store and retrieve participants' scores easily, they were asked to indicate their results in the Google form at the beginning of Task 4, as well as to provide general background information: age, level of education, and all the languages they speak (including the mother tongue).

The email was written in English in order to integrate the study into participants' learning process; basic vocabulary and structures accessible for learners at A2-B2 proficiency levels were used. Participants were advised to contact the researcher in case of any problem or doubt. For 3 out of 4 tasks (Tasks 1, 2, 4), both Spanish and Catalan versions were available,

so that the subjects could choose the language they were more comfortable with. The decision to provide the instruction in Spanish/Catalan rather than in English was an attempt to avert unnecessary misunderstanding and misinterpretation that could have been difficult to address in an online data collection setting. The v_YesNo test (Task 3) was available only in English and could not be modified. It was not considered a problem as it contained very short, simple instructions.

At the end of the email participants were notified about Session 2. It was specified that after the completion of the first session they would receive an email with the materials for the second one.

Participants performed the tasks autonomously on the device of their choice (a desktop computer or a laptop, but not a tablet or a phone) at a convenient time. We did not gather more detailed information on the setting learners chose to do the assigned experimental tasks, assuming they would be able to decide on the appropriate environment.

On completion of the first session, participants were assigned to one of the 4 genre groups based on their scores and their age. As the study followed between-subject design, a great effort was made to ensure groups comparability. The distribution of the key parameters across the groups can be seen below in Tables 5-10.

Table 5. Age across the genre groups

Genre group	N	Mean (SD)	Min	Max
Genre 1: sitcom	28	31.18 (16)	17	73
Genre 2: TED talks	29	31.97 (15)	17	60
Genre 3: drama	28	31.86 (15.3)	18	67
Genre 4: action	29	31.59 (16)	18	67

Table 6. Vocabulary size across the genre groups

Genre group	N	Mean (SD)	Min	Max
Genre 1: sitcom	28	4713.57 (1236)	2249	7650
Genre 2: TED talks	29	4796.17 (1120.8)	2550	8078
Genre 3: drama	28	4742.36 (1305.8)	2009	7617
Genre 4: action	29	4745.75 (1151)	2600	7112

Table 7. Pre-test scores across the genre groups

Genre group	N	Mean (SD)	Min	Max
Genre 1: sitcom	28	15.21 (3.7)	9	24
Genre 2: TED talks	29	15.55 (4)	11	27
Genre 3: drama	28	15.57 (4.4)	9	23
Genre 4: action	29	15.76 (3.8)	10	28

Table 8. Statistical learning measure 1 across the genre groups

Genre group	N	Mean (SD)	Min	Max
Genre 1: sitcom	27	4.08 (1.8)	0	8
Genre 2: TED talks	28	3.89 (2)	0	7
Genre 3: drama	27	3.67 (1.7)	0	7
Genre 4: action	29	3.9 (1.4)	1	6

Table 9. Statistical learning measure 2 across the genre groups

Genre group	N	Mean (SD)	Min	Max
Genre 1: sitcom	27	4.58 (1.9)	0	7
Genre 2: TED talks	28	4.89 (1.8)	1	8
Genre 3: drama	27	4.44 (2.2)	0	8
Genre 4: action	29	4.83 (2)	1	8

Table 10. Metalinguistic awareness across the genre groups

Genre group	N	Mean	Min	Max
Genre 1: sitcom	28	52.3 (9.4)	33	71
Genre 2: TED talks	28	51.8 (7.9)	32	66
Genre 3: drama	27	48.7 (7.8)	32	64
Genre 4: action	29	50.5 (7.4)	35	66

The participants were then sent an email with the materials for Session 2. It followed the same principles as the first one and contained a brief description of the session, links to the video sequence of the assigned genre and the post-test. The videos were presented as a traditional viewing comprehension activity and participants were told to stay focused during watching to be able to answer the questions that followed. As the post-test was exactly the same as the pre-test they already took, the subjects were notified not to be alarmed if the test seemed familiar. It was also emphasized that it had to be taken right after watching the videos. At the beginning of the post-test learners were asked to rank their viewing experience on a scale from 1 ("I didn't like the videos at all") to 5 ("I liked the videos a lot"). As this question was added after an insightful comment on a conference presentation of the preliminary results of the study, the data about video enjoyment is available only for 87 participants.

Session 2 was carried out under the same conditions that Session 1, autonomously on the device of participants' choice.

Initially, an attempt was made to maintain the exact one-week interval between the two sessions. It was considered optimal to prevent testing effect without lessening participants' engagement. The first 40 participants received an indication of the specific dates when they were supposed to do the second session. However, it proved to be unfeasible for the majority of them due to other commitments of personal or professional nature as well the structure of the language courses they were enrolled in. So instead, the rest of the subjects were allowed to complete Session 2 outside of the indicated time interval, and the analysis of the differences in time between the sessions was carried out to ensure it does not affect the results. The detailed information about time distribution overall and by group can be found in Table 11 below.

Table 11. Time between sessions per genre group

	Mean (SD)	Min	Max
Genre 1: sitcom	9.8 (4.6)	4	20
Genre 2: TED talks	11.5 (6)	1	32
Genre 3: drama	13.1 (10)	5	55
Genre 4: action	10.9 (6.5)	2	29
Overall	11.33 (7.1)	1	55

Although a slight variation in time between sessions across the genres can be observed, it was not statistically significant, as is shown in the preliminary analysis section below.

2.10. Preliminary analysis

This section describes the preliminary steps that were taken to explore the data before answering the main research questions.

To begin with, the data was reviewed to decide on the final sample. Overall, data was collected from 114 participants, however, it was incomplete for 6 of them. Besides the three participants mentioned earlier, who were not able to complete the statistical learning ability task, one participant did not report their vocabulary size score, and two participants submitted incomplete responses for the metalinguistic awareness test. These learners were excluded from the analysis, and the resulting sample was 108 participants.

Then the statistical analysis of the data was carried out, using IBM SPSS Statistics, versions 25.0 and 27.0.

2.10.1. Distribution of independent variables across genre groups

First of all, the assumed comparability of four experimental groups across the key parameters was statistically tested. One-way independent-samples analysis of variance (ANOVA) confirmed that neither participants' metalinguistic awareness (F (3, 104) = 1.04, p = .377) nor vocabulary size (F (3, 104) = 1.04, p = .377) differed significantly between the genre groups. Since the rest of the variables (age, the pre-test scores, both measures statistical learning ability) did not follow a normal distribution, a non-parametric equivalent of ANOVA, a Kruskal-Wallis test, was used for each of them. It was further confirmed that none of the variables were distributed significantly differently across the genre groups: $\chi 2$ (3, N = 108) = 1.5, p = .685 for the pre-test scores, $\chi 2$ (3, N = 108) = .30, p = .962 for age, $\chi 2$ (3, N = 108) = .55, p = .907 for statistical learning measure 1, $\chi 2$ (3, N = 108) = .422, p = .936 for statistical learning measure 2.

Therefore, it was established that the independent variables were balanced across the genre groups, making any between-group comparisons valid. The contribution of other possible intervening variables that could not be controlled for in the study is further discussed below.

2.10.2. Intervening variables: time between the sessions

Since it was impossible to control for the exact amount of time between the two experimental sessions across participants, and these differences meant that some learners could remember the pre-test better than others, this variable was analysed in more detail to test the possibility of its effect on the learning outcomes.

First, the differences in time between the sessions were explored at the group level. A Kruskal-Wallis test showed that the four genre groups did not differ in this parameter at a

statistically significant level: $\chi 2$ (3, N = 108) = 2.485, p = .478, with Md = 9 for sitcom, Md = 12 for TED talks, Md = 9 for drama, Md = 9 for action.

Secondly, a Spearman's correlation revealed no relationship between the time between the sessions and learning gains, $r(106) = -.075$, p = .442.

It was therefore concluded that the varied interval between the two experimental sessions did not impact learning and this variable was excluded from the further analysis.

2.10.3. Intervening variables: previously seen episodes

Overall, participants did not report high familiarity with the videos in the treatment. There were only 3 learners in each genre group who had seen some of the presented videos before. The extent of their familiarity is described in Table 12.

Table 12. The ratio of familiar videos for each of the participants who reported familiarity

	Participant 1	Participant 2	Participant 3
Sitcom	12/12	7/12	3/12
TED talks	1/12	2/12	1/12
Drama	3/12	3/12	3/12
Action	3/12	4/12	1/12

As it can be seen, there were two participants who were familiar with all or most of the sitcom videos, with the rest of the genres being largely unfamiliar to the learners. All subsequent analysis was run with and without these six subjects, and since the exclusion did not have any impact on the results they were kept in the sample.

2.10.4. Intervening variables: enjoyment

After the treatment participants were asked to rate their enjoyment of the videos on the scale from rom 1 ("I didn't like them at all") to 5 ("I liked them a lot"). As this question was added when data collection was already in progress, the data is available for 87 learners of the total sample and for 83 learners of the resulting filtered sample. The distribution of the scores from the filtered sample is reported in Table 13 below.

Table 13. Participants' enjoyment of the videos across the genre groups

Genre group	N	Mean (SD)	Min	Max
Sitcom	20	4.15 (1.3)	1	5
TED talks	21	3.76 (0.9)	2	5
Drama	20	3.25 (0.9)	1	5
Action	22	3.55 (1.2)	1	5

A Kruskal-Wallis test confirmed significant differences in enjoyment of the videos from different genres, $\chi 2$ (3, N = 83) = 9.7, p = .022. Pairwise comparisons revealed that these differences were significant between sitcom and drama (p = .022, *Md* =5 and *Md* =3, respectively) as well as between sitcom and action (p = .048, *Md* =5 and *Md* =4, respectively), but not between sitcom and TED talks (p = .098, *Md* =5 and *Md* =4), TED talks and drama (p = .145, *Md* =4 and *Md* =3), TED talks and action (p = .756, *Md* =4 and *Md* =4) or drama and action (p = .244, *Md* =3 and *Md* =4). This suggests that overall, sitcom was the most enjoyable genre, followed by TED talks, action, and drama, with not all the differences, however, reaching a significant level.

Table 14 shows in more detail how participants' ratings varied within each genre group. If we regard answers "1" and "2" as indicators of low enjoyment, and answers "4" and "5" as indicators of high enjoyment, the ratings confirm that sitcom was liked the most (16 high vs. 2 low) and drama the least (8 high vs.3 low). TED talks genre seems to be the most "neutral" (11 high, 1 low, 9 medium), while action is the genre that received the most polarised ratings (13 high vs. 4 low).

It is noteworthy that these differences could be interpreted mainly as trends, as none of the genres were consistently disliked and individual preferences seemed to have considerable weight in the proposed "enjoyability" quality, as even the sitcom, appearing as the most entertaining here, was rated as not enjoyable at all by two participants.

Table 14. Frequencies of each scale point across the genre

Point on scale	Sitcom	TED talks	Drama	Action
1	2	1	1	2
2	0	0	2	2
3	2	9	9	5
4	5	5	7	8
5	11	6	1	5

However, there was no correlation between participants' reported enjoyment of the videos and their learning gains, r (81) = -.078, p = .481.

Chapter 3. Results

This chapter reports the results of the present study. It contains five main sections, which sequentially cover the description of learning conditions under the four genres used in the study, examination of the factors affecting the learning of formulaic sequences from captioned videos as well as any possible interactions between the context-related and learner related variables (research questions 1, 2, 3, 4), and exploration of metalinguistic awareness across formulaic types and test components (research question 5). The chapter is concluded with a brief summary of the findings.

3.1. Learning conditions under different genres

Before examining possible effects of different learning conditions associated with different genres, these conditions need to be defined. The detailed description of the conditions is provided in the methodology chapter (Table 4) and here the differences are analyzed and summarized.

First of all, variations between genres were present at the level of vocabulary demands. To pass the threshold of 95% coverage, considered as the lower boundary for adequate comprehension, learners needed to know 2000 most frequent word families (plus marginal words and proper nouns) for drama (97.1%), sitcom (96.83%), and action (95.2%), while 3000 most frequent word families were needed for the TED talks genre (95.62%). Similarly to Webb and Rodgers (2009a), the differences were more pronounced at the level of 98% coverage. Four thousand word families (4000), plus marginal words and proper nouns, provided 98.1% coverage for sitcom and 98.24% for drama, while 5000 and 6000 word

families were needed to reach 98.45% and 98.02% coverage for action and TED talks, respectively.

The same pattern emerged from the lexical complexity analysis, which showed that the overall percentage of low-frequency words was the smallest in drama, followed by sitcom, action, and TED talks. Thus, both lexical measures suggest that TED talks were the most challenging genre in terms of vocabulary, with action as a close second, while drama and sitcom were the easiest ones and were quite comparable in their vocabulary load.

Identified genre-related differences in learning conditions were not limited to the lexical aspect. Such elements of L2 videos as action type, plot organization, type of narration, contextual clues, and the required background information were considered equally relevant for learning, and the genres were compared in terms of these elements as well. It is noteworthy, however, that this comparison is of more speculative character since these features are not quantified here or compared statistically.

In terms of action type, we can theorize that the action genre can be the most challenging one since it contains not only action-filled pauses, but also action scenes competing with narration. Therefore, participants' attention could be potentially divided between dialogue and image, which cannot be considered optimal for language learning. As for the other three genres, both sitcom and drama contain complementary action, and its effect on learning conditions cannot be easily predicted; while TED talks have no action involved, which deprive learners of some imagery support, at the same time allowing them to focus on the narration without any distractions.

The contribution of plot organization to learning conditions is defined through the ease or difficulty of following an individual episode of a genre. It therefore indirectly determines if an individual episode of a program can be effectively integrated in multimedia instruction. TED talks appear to be the easiest genre in this respect since most of the talks contain full stories and did not overlap. It is followed by sitcom, with a certain degree of the overlap between the episodes. Drama and action have the highest degree of the overlap, with considerable familiarity of previous episodes needed in order to follow the story.

In terms of required background information, sitcom and TED talks seem to impose comparable demands on learner, while action appears to have the highest requirements, followed by drama.

Number of contextual clues available for learners' interpretation of pragmatic formulaic sequences seems to be connected with a narration type, as it is higher for the genres that involve a dialogue (sitcom, drama, action), while in TED talks learners can rely only on speaker's facial expressions, gestures, and intonation in interpreting the meaning of pragmatic formulas.

To conclude, it was shown that the four selected genres indeed varied in the provided learning conditions, with sitcom and drama possibly the least demanding genres overall, and TED talks and action potentially more challenging across a number of aspects. It is tentatively suggested that action could be the most difficult genre to follow, due to a high number of elements it contains and the fact that action scenes might be competing with narration for attentional resources. Sitcom seems to have slight advantage over drama, since there is less overlap between the episodes, suggesting that individual episodes of sitcom can be used as a

part of multimedia instruction. Having established that, we proceed to examine the data on learning in more detail to determine whether our theoretical predictions hold true.

3.2. General impact of the treatment: pre- post-test comparison

Since one of the objectives of the study was to examine factors affecting learning, it first had to be proved that learning actually took place and resulted from the experimental treatment. To determine whether participants' knowledge of the target formulaic sequences improved after watching the videos, a series of related-samples tests was administered to compare the pre- and post-test scores for each genre group. The scores came from the Vocabulary Knowledge Scale (VKS; Wesche & Paribakht, 1996), and for each target item they ranged from 1 (*"I have never seen this expression before"*) to 5 (*"I know this expression and I can use it the context"*; correct translation and example in the context provided). Therefore, the possible total score for each participant per test had a minimum of 6 and a maximum of 30.

If the scores were normally distributed, a paired-samples t-test was used (sitcom, action), otherwise a Wilcoxon Signed Rank test was run (TED talks, drama). Three out of four genre groups demonstrated a higher knowledge of the target items at the post-test compared to the pre-test: sitcom, $t(26) = -6.1$, $p < .001$ ($M = 15.4$, $SD = 3.9$ at the pre-test; $M = 17.7$, $SD = 4.3$ at the post-test; large effect size $drm = 1.3$), TED talks, $z = -3.2$, $p = .001$ ($Md = 14$, $n = 27$ at the pre-test, $Md = 16$, $n = 27$ at the post-test; moderate effect size r = .43), and drama, $z = -3.1$, $p = .002$ ($Md = 14$, $n = 26$ at the pre-test, $Md = 17.5$, $n = 26$ at the post-test; moderate effect size r = .43). There was no significant improvement between the pre-test ($M = 16.29$, $SD = 3.9$) and post-test scores ($M = 17.32$, $SD = 3.7$) for action genre, $t(27) = -1.93$, $p = .064$.

So as to ensure that the observed learning effect could be attributed to the treatment, participants knowledge of the distractors was compared at the pre- and post-tests. Since they did not appear in the videos, any significant improvement would mean a strong test effect or exposure outside of the study (classroom, dictionaries, asking a friend, etc.). Paired-samples t-tests showed that learners' knowledge of the distractors did not differ significantly ($t(26) = -.13$, $p = .899$) between the pre- ($M = 31.81$, $SD = 4.7$) and post-test ($M = 31.89$, $SD = 6$) for the sitcom, neither it did for TED talks ($t(26) = .071$, $p = .944$, $M = 31.33$, $SD = 5.4$ at the pre-test, $M = 31.30$, $SD = 5$ at the post-test), neither it did for drama ($t(25) = -.92$, $p = .363$, $M = 32$, $SD = 6.4$ at the pre-test, $M = 32.38$, $SD = 6.6$ at the post-test), neither it did for action ($t(27) = -1.6$, $p = .111$, $M = 32.82$, $SD = 6$ at the pre-test, $M = 33.43$, $SD = 6.8$ at the post-test). It can thus be inferred that if any learning occurred, it was due to the exposure to the captioned videos during the study.

In sum, it was demonstrated that statistically significant learning of the target items occurred in the sitcom, TED talks, and drama groups, but not in the action one. Thus, a variable of learning gains was computed as the difference between the pre- and post-test scores for each participant and was used as the measure of learning further on. Table 15 displays the descriptive statistics for the variable for each of the genre groups. There was no significant improvement of participants' knowledge of the distractors across the genres. Therefore, the validity of captioned videos as a tool promoting incidental learning of formulaic sequences was confirmed.

Table 15. Gains across the genre groups

Genre group	N	Mean (SD)	Median	Min	Max
Sitcom	27	2.26 (1.9)	2	0	7
TED talks	27	1.78 (2.4)	1	-3	7
Drama	26	1.54 (2.2)	1	-3	7
Action	28	1.04 (2.8)	1.5	-7	6

3.3. Research questions 1-4: factors affecting learning and interactions

This section seeks to answer research questions 1, 2, 3, and 4, and namely to establish whether the observed learning of formulaic sequences from L2 videos was affected by the genre of the videos, participants' metalinguistic awareness and their statistical learning ability, as well as to investigate any possible interactions between learner-related and context-related factors. Although age, vocabulary size, and the pre-test scores were not explicitly targeted in the research questions, their contribution to learning process could be important, based on previous research and high heterogeneity of our sample; it is therefore addressed here as well. For the statistical learning measure, the two types of trials (words vs. part-words and words vs. non-words) were analyzed separately as statistical learning measure 1 and statistical learning measure 2, correspondingly. The reason for this separation is that they might probe into slightly different aspects of this ability, with words vs. part-words trials measuring "pure" statistical leaning, and words vs. non-words trials involving working memory component as well.

To explore how the selected set of independent variables impacts learning outcomes and to identify their relative weight, a series of general linear models (GLM) was run. First, to examine simple effects of all independent variables, a model was fit with gains as a dependent variable, genre as a categorical predictor, and participants' age, vocabulary size, metalinguistic awareness, statistical learning ability (two separate scores for two trial types), and the pre-test scores as continuous predictors. Since the original scale of vocabulary size was considerably larger than the rest of the variable, adjusted vocabulary size was computed by dividing the original score by 1000 (e.g., 5600 became 5.6). Collinearity diagnostics was run to ensure that none of the continuous predictors were significantly correlated, as the opposite could result in a biased model. Variance Inflation Factor (VIF) was below 3 for all the variables (see Appendix E), which is considered as the upper boundary for acceptable multicollinearity (Kock & Lynn, 2012). Thus, all the variables were included in the same model.

The overall model was significant, F (9,98) = 2.5, p = .012, $\eta p2$ = .187. It showed no significant main effect of genre, F (3,98) = .72, p = .543, $\eta p2$ = .022, statistical learning measure 1 (words vs. partwords), F (1,98) = .061, p = .806, $\eta p2$ = .001, statistical learning measure 2 (words vs. nonwords), F (1,98) = .282, p = .596, $\eta p2$ = .003, adjusted vocabulary size, F (1,98) = 1.15, p = .287, $\eta p2$ = .012, and age, F (1,98) = 3.38, p = .053, $\eta p2$ = .038. Metalinguistic awareness (F (1,98) = 7.9, p = .006, $\eta p2$ = .075) and the pre-test scores (F (1,98) = 12.165, p = .001, $\eta p2$ = .110) emerged as significant predictors of learning gains. Table 16 [17] shows estimated means for the between-groups factor (genre).

[17] Continuous within-group predictors appearing in the model are evaluated at the following values: Pre-test = 15,62, Age = 32,00, MetAwar = 51,0185.

Table 16. Estimated means of learning gains by genre

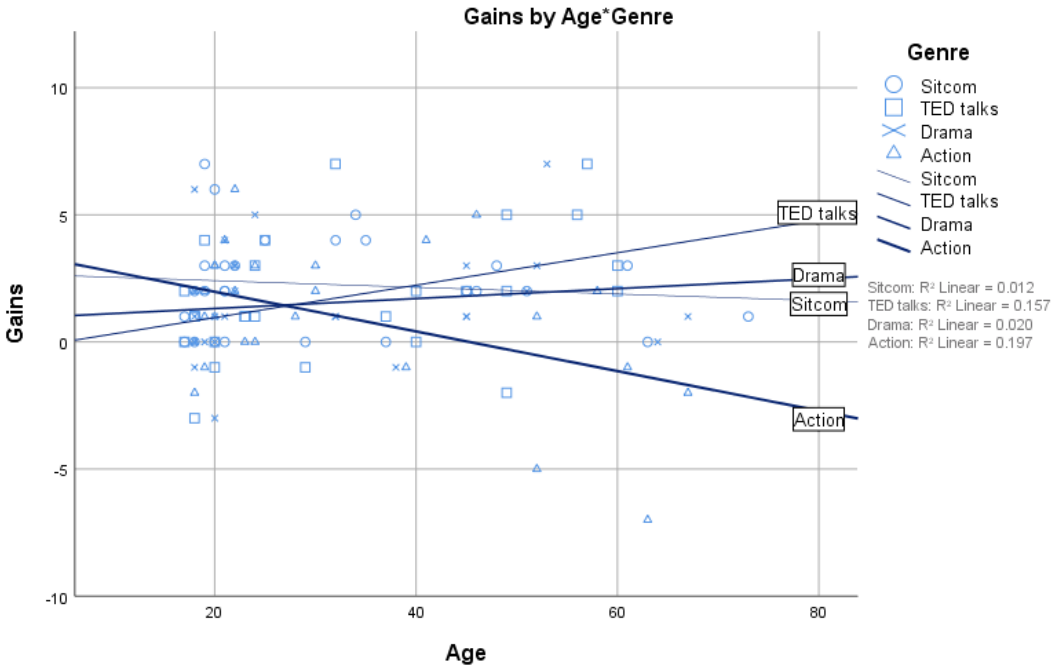
Genre group	Mean (SE)	95% CI
Sitcom	2.112 (.420)	[1.279 – 2.945]
TED talks	1.572 (.420)	[.738 – 2.405]
Drama	1.718 (430)	[.865 – 2.570]
Action	1.178 (412)	[.361 – 1.996]

With the purpose of investigating the individual contribution of metalinguistic awareness components, the scores for each task of the metalinguistic awareness test were loaded into the same model individually. The model returned a significant main effect of learners' performance at Task 1 (identifiability in the text), F (1,94) = 5.1, p = .026, $\eta p2$ = .051, while their scores from Task 2 (identifiability in the list, F (1,94) = 1.4, p = .241, $\eta p2$ = .015) Task 3 (meaning-shifts, F (1,94) = 1.16, p = .284, $\eta p2$ = .012), Task 4 (fixedness, F (1,94) = .858, p = .357, $\eta p2$ = .009), and Task 5 (compositionality, F (1,94) = .380, p = .539, $\eta p2$ = .004) were not significant; the significance of the other predictors was unaffected. However, although participants' ability to identify formulaic sequences in the text appeared to be the strongest component of metalinguistic awareness, its individual effect was not that stable across the models and was lowering the overall explanatory power of a model. Therefore, the global metalinguistic awareness score was used for the further analysis as a more comprehensive measure.

In order to explore the data further, a full model was built following a backward stepwise regression method. We started with a saturated model that included all predictors as well as

two-way and three-way interactions between them, and then eliminated those that did not significantly improve the model one by one. Akaike information criterion (AIC) model selection (Cavanaugh & Neath, 2019) was used to determine the best fitting model. The resulting model was significant (F (9,98) = 3.55, p = <.001, $\eta p2$ = .246), and confirmed that the pre-test scores were the strongest single-term predictor of learning gains (F (1,98) = 8.4, p = .005, $\eta p2$ = .079), followed by metalinguistic awareness (F (1,98) = 8.12, p = .005, $\eta p2$ = .079), with higher pre-test scores resulting in lower gains (B = -.173, t = -2.9, p = .005) and metalinguistic awareness having a facilitatory effect on learning (B = .082, t = 2.85, p = .005). A significant interaction between genre and age could be observed as well (F (3,98) = 3.2, p = .026, $\eta p2$ = .09). The interaction (visualized in Figure 5) indicates that while there was no to little impact of age on learning in sitcom and drama groups, participants' learning gains from watching the TED talks increased with their age, and the opposite trend emerged for the action videos, where the older the learners were, the less they improved their knowledge of the target items after the treatment. Since the final model included a significant interaction term, the simple effects of genre and age are not reported.

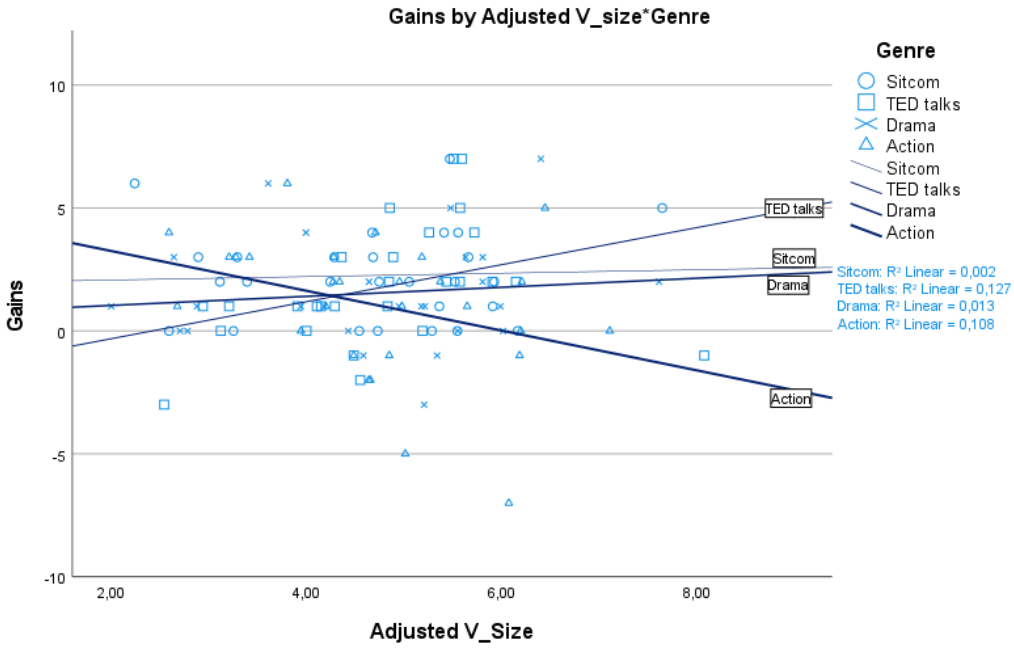
Figure 5. Interaction between genre and age



Although the interaction between genre and adjusted vocabulary size was not significant (F (3,97) = 2.24, p = .087, $\eta p2$ = .065), it is interesting to mention that a very similar pattern could be observed for this variable as well (Figure 6).

To examine whether the negative impact of high performance at the pre-test could be interpreted as the ceiling effect, the same model was run with the data filtered to include only the participants who scored 18 or lower (out of 30) before the treatment. The model sill returned a significant effect of the pre-test scores (F (1,70) = 5.4, p = .023, $\eta p2$ = .072), and the other predictors were not affected either. It could be therefore inferred that in general, if participants who knew the target items better before the treatment learned less from the treatment, it is not due to the lack of space for improvement and should be explained in other terms.

Figure 6. Interaction between genre and adjusted vocabulary size



Finally, it was ensured that the data met the assumptions necessary for a reliable interpretation of the model.

Appendix F demonstrates that standardized residuals were normally distributed. It means that a linear model fitted the data well, despite the non-normal distribution of the gains variable.

Levene's test showed that the assumption of homogeneity of variance across groups was not violated (F (3,104) = .413, p = .744), and Breusch–Pagan test confirmed homoscedasticity (consistency along the dependent variable) of the residuals ($\chi 2$ (9, N = 108) = 9.2, p = .416).

Further on, to check if there were any participants whose behavior deviated significantly from the established patterns thus biasing the model, Cook's Distance values were calculated. Since all the values were below 1 (max = .16), there were no influential cases interfering with

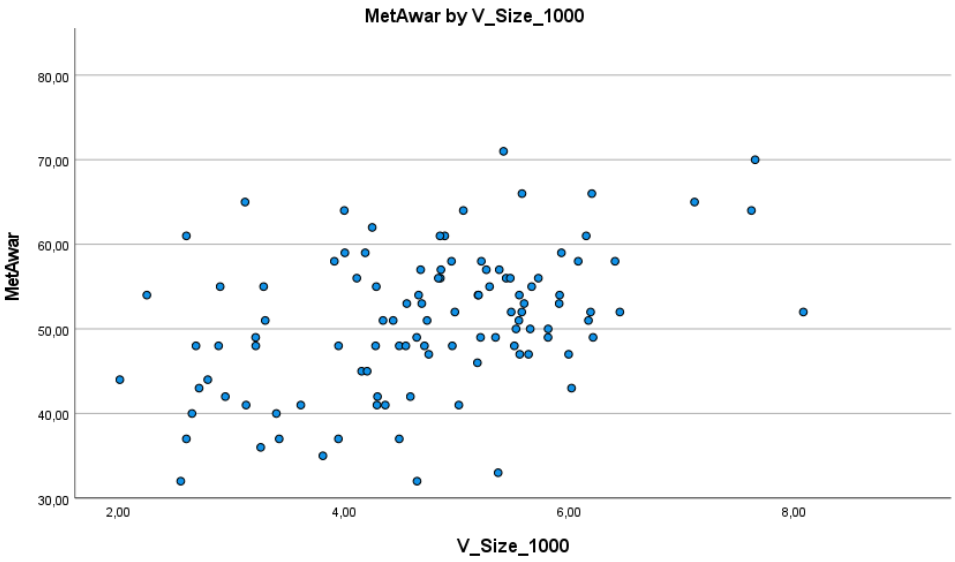
the results (Cook, 2011). It was therefore established that the model interpretations could be trusted.

3.3.1. Exploring metalinguistic awareness

In this section we intend to answer research question 5 and explore in more detail participants' metalinguistic awareness.

First of all, the relationship between metalinguistic awareness scores and adjusted vocabulary size was examined. Pearson's correlation revealed that there was a significant positive relationship between the two variables, but it was rather weak (r (108) = .41, p<.001). If we reduced the variability in vocabulary size scores by filtering the data to exclude participants that knew less than 3500 words (3.5 in the adjected score) and more than 7000 (7 in the adjusted score), the strength of the correlation decreased to a very low level (r (83) = .25, p = .021). High variation in metalinguistic awareness scores can be also seen in Figure 7 below.

Figure 7. Distribution of metalinguistic awareness scores across adjusted vocabulary sizes



It can be therefore concluded that although a certain relationship between metalinguistic awareness and proficiency (as measured by vocabulary size) could be observed, it was not as straightforward as it could have been expected.

3.4. Research question 5: metalinguistic awareness across formulaic types

To explore how participants' metalinguistic awareness varied across formulaic types, a series of generalized mixed models was run for each component of the test (identifiability in the text, identifiability in the list, lexical fixedness, and compositionality[18]). Each model was based on simple binary logistic regression, since the dependent variable was participants' answer for each formulaic item in the test, which was coded as "0" (incorrect) and "1" (correct). The random effect included subject, and the predictor was formulaic type. The results of the models are reported below.

3.4.1. Identifiability in the text

The model correctly classified 75.8% of cases and returned a significant effect of formulaic type, F (5, 2010) = 41.47, $p < .001$. Estimated means per type are displayed in Table 17 below, and Figure 8 at the end of the section provides visual representation of the answer pattern by type.

[18] Since Task 3 (meaning-shift component of the test) included only one formulaic type, it was not analyzed here.

Table 17. Estimated means of participants' answers in Task 1 by formulaic type

Type	Mean (SE)	95% CI
Collocation	.284 (.030)	[.229 – .347]
Idiom	.512 (.035)	[.443 – .580]
Pragmatic	.199 (.025)	[.153 – .253]
Binomial	.166 (.023)	[.126 – .217]
Phrasal verb	.551 (.035)	[.482 – .619]
Lexical bundle	.163 (.023)	[.123 – .213]

3.4.2. Identifiability in the list

The model correctly classified 67.3% of cases and returned a significant effect of formulaic type, F (5, 1338) = 22.26, $p < .001$. Estimated means per type are displayed in Table 18 below, and Figure 9 at the end of the section provides visual representation of the answer pattern by type.

Table 18. Estimated means of participants' answers in Task 2 by formulaic type

Type	Mean (SE)	95% CI
Collocation	.572 (.035)	[.503 – .639]
Idiom	.459 (.035)	[.392 – .528]
Pragmatic	.505 (.035)	[.436 – .573]
Binomial	.324 (.033)	[.263 – .391]
Phrasal verb	.833 (.026)	[.777– .877]
Lexical bundle	.509 (.035)	[.440 – .577]

3.4.3. Lexical fixedness

The model correctly classified 86.8% of cases and returned a significant effect of formulaic type, F (3, 718) = 26.24, $p < .001$. Estimated means per type are displayed in Table 19, and Figure 10 at the end of the section provides visual representation of the answer pattern by type.

Table 19. Estimated means of participants' answers in Task 4 by formulaic type

Type	Mean (SE)	95% CI
Collocation	.920 (.019)	[.873 – .951]
Idiom	.858 (.028)	[.794 – .904]
Pragmatic	.896 (.023)	[.842 – .933]
Binomial	.521 (.052)	[.419 – .622]

3.4.4. Compositionality

The model correctly classified 78% of cases and returned a significant effect of formulaic type, F (3, 1332) = 57.3, $p < .001$. Estimated means per type are displayed in Table 20 below, and Figure 11 at the end of the section provides visual representation of the answer pattern by type.

Table 20. Estimated means of participants' answers in Task 5 by formulaic type

Type	Mean (SE)	95% CI
Collocation	.412 (.039)	[.338 – .490]
Idiom	.898 (.019)	[.855 – .929]
Pragmatic	.415 (.039)	[.340 – .493]
Phrasal verb	.423 (.039)	[.348– .501]

Figure 8. Scores across types in Task 1

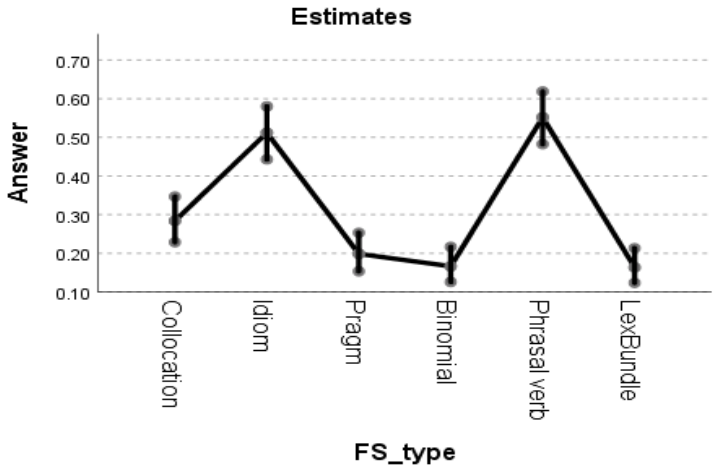


Figure 9. Scores across types in Task 2

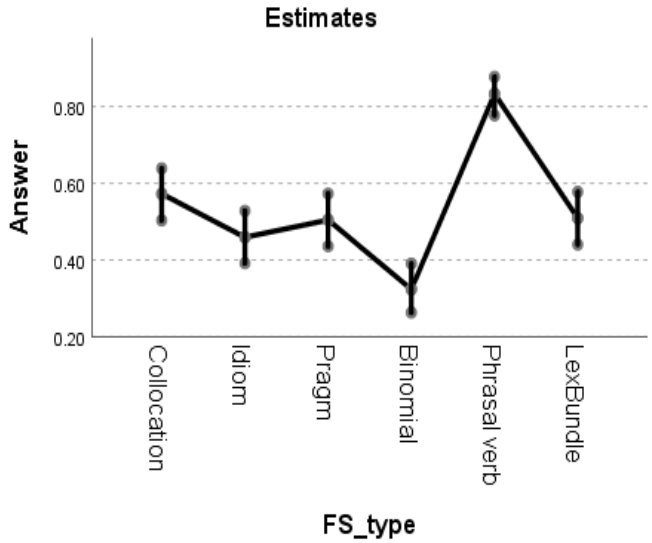


Figure 10. Scores across types in Task 4

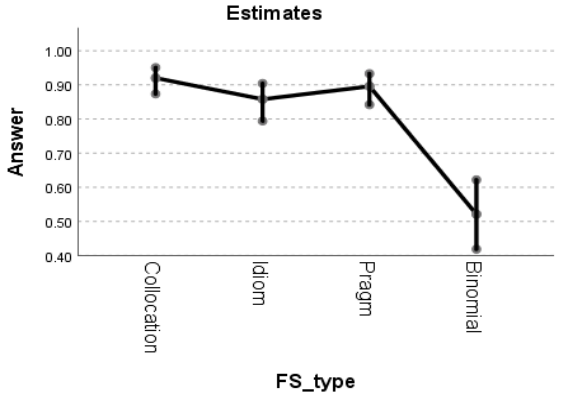
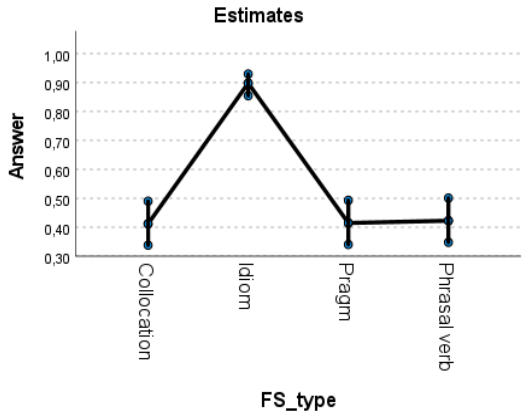


Figure 11. Scores across types in Task 5

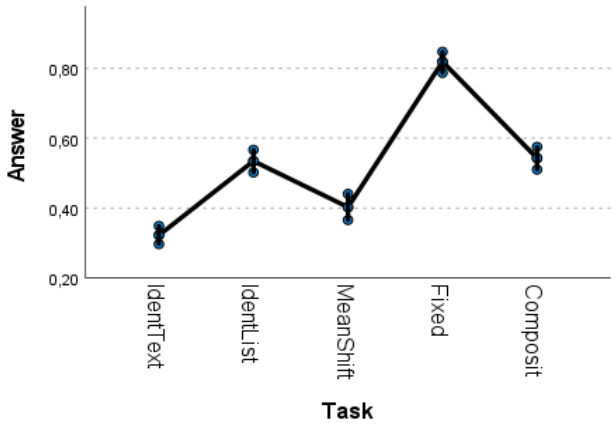


3.5. Research question 5: metalinguistic awareness across components

Finally, another model was built to assess participants' performance across the five test components. The model correctly classified 64.8% of cases and returned a significant effect of formulaic type, F (4, 6281) = 128.5, p < .001. Figure 12 shows how learners' scores

fluctuated across the tasks that probed into sensitivity to identifiability (text and list), meaning shifts, fixedness, and compositionality, respectively. The only insignificant difference was between Task 2 (identifiability in the list) and Task 5 (compositionality), where Task 2 did not appear to be more challenging than Task 5, B = -.034, t = -.429, p = .668.

Figure 12. Participants performance across test components



3.6. Summary of the results

1. Learning conditions under the four selected genres were identified and compared, revealing certain variation across the genres.
2. Captioned L2 videos were confirmed as a valid tool for promoting incidental learning, with three out of four genre groups improving their knowledge of the target items significantly from pre- to post-test. The action genre did not prove to be beneficial for learning of pragmatic formulaic sequences.

3. The context-related factor of genre interacted with participants' age (at the significant level) and vocabulary size (not significant). While sitcom and drama were generally beneficial for learners of different ages and vocabulary sizes, the genre of TED talks was more advantageous for older participants and those who had a higher vocabulary size, and the opposite trend was present for the action genre. Due to the interactions, the main effect of genre could not be reliably measured, but the interactions provided interesting insight into the suitability of different genres for participants with different background characteristics.
4. From the learner-related factors, participants' pre-test scores and metalinguistic awareness emerged as significant predictors of learning outcome, with higher pre-test scores leading to lower gains and higher metalinguistic awareness leading to higher gains. The impact of the pre-test could not be attributed to the mere ceiling effect. Statistical learning ability did not play any role in the learning of pragmatic formulaic sequences.
5. Learners' metalinguistic awareness varied across the formulaic types and formulaic aspects.

Chapter 4: Discussion

This chapter is devoted to the consolidation and interpretation of the findings of the dissertation. It includes 10 sections that are organized as follows: first, the selected learning context of multimedia instruction and its effectiveness for acquisition of formulaic sequences are reviewed in light of the obtained results; then context- and learner-related factors that were brought into focus are discussed as well as their individual contributions to the observed

learning outcomes; finally, learners' metalinguistic awareness of various formulaic types and main aspects of formulaicity is commented on. The last two sections contain limitations, conclusions, and pedagogical implications.

4.1. Learning formulaic sequences from audio-visual input

The present study aimed at exploring the learning of pragmatic formulaic sequences from captioned videos and the factors that might affect this process, with these factors relating to either the context or the learners. During the study, Catalan/Spanish learners of English were exposed to approximately 30 minutes of captioned L2 videos. In line with Frumuselu et al. (2015), Moskvina (2017), Majuddin et al. (2021), and Vu et al. (2023), this exposure led to overall significant improvement in participants' knowledge of the formulaic sequences that were present in the audio-visual input, as measured by intake. Thus, our results further confirm the potential of captioned videos to promote incidental learning of formulaic language, showing that the beneficial effects could be extended to pragmatic formulaic sequences, since previous research has not consistently brought this formulaic type into focus (with the exception of Moskvina, 2017). It is an encouraging finding since this formulaic type is highly context-dependent and learners need extensive exposure to the contexts where these sequences occur naturally to be able to form accurate intuitions about their use. The obtained results suggest that this exposure could be efficiently provided by authentic captioned videos in the target language. However, despite the overall beneficial effect of multimedia instruction and similarly to the previous studies, the learning outcomes in our research varied according to a number of factors.

4.2. Learning conditions under different genres

Since the study involved four different treatment groups that were exposed to the videos of four different TV genres (sitcom, TED talks, drama, and action), with the target formulaic items being the same across the groups, we were able to compare the learning conditions created by each of the genres as well as the actual impact of these conditions on the learning of formulaic sequences.

The results build on and extend the findings from the corpus studies by Webb and Rodgers (2009a, 2009b), demonstrating variation in the learning conditions provided by different genres. In line with previous research, the genres used in the study differed in the lexical demands that they imposed on learners. The lexical coverage data suggested that the vocabulary demands were the lowest for the genres of drama and sitcom, followed by action and TED talks. Although we cannot make large-scale generalizations about the whole genre based on a 30-minute excerpt, our findings seem to exhibit the same trends as Webb and Rodgers (2009b) and therefore can be used as an indicator of the overall vocabulary demands of the genres used in the study.

While previous corpus studies focused exclusively on the lexical aspect of the learning conditions provided by various L2 movies and TV programs, results in this dissertation have demonstrated that the genre-related differences in these conditions go beyond vocabulary demands and are equally defined by the peculiarities in plot and narration organization, image and its components. This is partially in line with Rodgers (2018), who found that imagery support varied between documentary and narrative programs; however, since the primary focus of this dissertation is formulaic language, and namely pragmatic formulas, the interpretation of imagery had to be adjusted accordingly. It is therefore suggested that in the

case of pragmatic formulaic sequences, the supporting role of image, advocated by Mayer's (2009) multimedia principle, lies in supplying learners with additional information about the communicative situation, e.g., gestures, intonation, facial expressions reflecting speaker's attitude, interlocutors' reaction, lead-up to the situation and other indicators that might help the viewer match the presented context with the corresponding experience from their life. It is further proposed here that imagery support of audio-visual input should be considered in terms of contextual clues.

The other non-language related aspects of L2 videos that can potentially influence multimedia learning include action type, plot organization, type of narration and the required background information. It is suggested that these aspects can cause variation in the amount of details the viewer needs to pay attention to in order to effectively follow the narration and as a result affect the working memory load of a genre. Therefore, according to Sweller's Cognitive Load Theory (1994), these factors are relevant to consider in the design of efficient multimedia instruction that would ensure optimal use of learners' working memory resources.

Based on the theoretical analysis of the learning conditions, sitcom appeared to be the most beneficial genre for language learning, followed by drama and TED talks, while action seems to be the most challenging one. The experimental testing of these assumptions is further discussed below.

4.3. The effect of genre on learning

Since the learning potential of different genres has not been extensively explored at the experimental level, it is not quite easy to place our findings into context. There is little overlap

between previous studies, as they dealt either with different types of target items, different set of genres or independent variables, or only one genre entirely. However, we will try to connect the pieces of the evidence obtained so far in an attempt to further our understanding of how genre might affect incidental learning from audio-visual input.

The results suggest that not all four genres used in the study (sitcom, TED talks, drama, and action) were equally beneficial for learning. While sitcom, TED talks, and drama groups improved learners' knowledge of the target formulaic sequences after the treatment, no significant changes in participants' performance occurred in the group that watched action videos. It shows that the genre of action is less advantageous for incidental learning of pragmatic formulaic sequences. This finding reflects the differences in the learning conditions discussed above. The action genre imposes high demands on learners' working memory, with a lot of movement on the screen and a high number of details necessary to follow the story. Therefore, pragmatic formulas are not prioritized by the viewer since they do not contribute to the development of the plot, and interpersonal relationships are not the main focus of the genre.

Heterogeneity of the sample did not allow for further robust claims about the overall differences between the genres. However, it let us explore the suitability of the chosen genres for various population groups. While age did not have a significant impact on learning in the drama and sitcom groups, it did so for TED talks and action. Older participants seem to benefit from watching TED talks more than the younger ones, and the opposite was the case for action, where the learning gains declined with participants' age. Once again, we believe that the learning conditions provided by the genres could account for this finding. TED talks are the least dynamic and engaging of the four selected genres, and they covered a variety of

global topics, such as healthcare, social stereotypes, school systems, etc. It is logical to suggest that these kinds of topics would be more appealing to 40-year-olds than late teenagers and university students, and therefore the older population would probably pay closer attention to the content. Another thing to remember here is that more mature students in our study were recruited from language schools, where they were attending regular classes in their free time. This might indicate a high level of motivation and, as a result, willingness to take advantage of the presented audio-visual materials regardless of the engagement value of the content, i.e., they are more likely to view this activity as an extra exercise for their language studies rather than casual "watching for pleasure". This could often be seen in their communication with the researcher before and after the experiment, as they expressed eagerness to participate and were curious about their results.

The action genre, as it has been mentioned above, is the most demanding genre of the four, with the learning conditions that it creates possibly overloading participants' working memory. It is therefore logical to suggest that this genre was even more challenging as participants' age increased, since working memory capacity is known to decline with age (Mattay et al., 2006). Besides, it is the most recent TV genre, and the younger generation might be more familiar with the superhero subculture, which would facilitate following the story and free attentional resources for incidental language learning.

A similar pattern was observed for the effect of vocabulary size across the genre groups. We are careful in the generalizations of this aspect, since a low number of participants per group did not allow the interaction to reach a significant level, however, it seems noteworthy. Unlike Puimège and Peters (2019a, 2020) and Majuddin et al. (2021), who reported a general facilitatory effect of vocabulary size on incidental learning of formulaic sequences, this

variable did not appear to have any significant impact on learning gains in the sitcom and drama groups. The reason for this difference in the results might lie in the nature of the target items across the studies. While the previous studies involved formulaic sequences comprised of individual words varying in frequency and potential familiarity to learners (e.g., "slippery slope" and "whisked away" in Majuddin et al., 2021; or "grist to the mill" and "sexual depravity" in Puimège and Peters, 2020), the formulaic sequences investigated in the present study consisted mainly of high-frequent words potentially already familiar to participants. Therefore, previous vocabulary knowledge might be more important for more lexically challenging sequences containing less frequent words, and in the case of formulaic sequences composed of familiar words, other learners' abilities might have more weight.

Similarly to the age variable, vocabulary size had more weight in the TED talks and action groups. Learners with a larger vocabulary size could benefit more from the TED talks treatment and less from the action one. The facilitating effect of previous vocabulary knowledge in the TED talk groups is not surprising, considering that this genre emerged as the most lexically demanding in the theoretical analysis of the genre-related learning conditions. Despite the aid of the introductions that appeared before each video in order to contextualize its content and facilitate comprehension, language learners might still have found TED talks to be the most challenging genre in terms of key vocabulary. That would lead them to prioritize words and word combinations necessary for understanding and ignore the rest. Since pragmatic formulas normally do not convey any meaning that would advance the plot but are rather used to modify the flow of communication, they might be easily ignored in this case. Additionally, this genre offers no image support for the content, so learners might have had to rely more heavily on the verbal component of the videos for

content interpretation. Suarez et al. (2021) commented on the relationship between vocabulary size and imagery support as well, in their case with the example of single words. The authors suggested that language learners relied less on their vocabulary size when watching a documentary than they did when watching sitcom, crime, or TED talk videos because the genre of documentary provides stronger support of the context through imagery, thus lessening the vocabulary burden of the genre. It can be inferred here that the opposite might be true for the TED talks in our study, with learners resorting to their vocabulary resources more in the absence of the imagery aid.

The vocabulary size effect had a completely opposite direction for the action group in our study, with learners who had a larger vocabulary size showing lower gains. This is partially in line with Vu et al. (2023), who found that prior vocabulary knowledge had a negative impact on incidental learning of collocations from audio-visual input. Although their research involved a different genre (adventure, not explicitly described in the study but identified as such by the IMDB website, an online film and TV database; IMDB, n.d.), a similar explanation could be extended for both studies: participants with a larger vocabulary size focused more on the plot development than the language aspect of the videos. This reiterates our statement about the challenging learning conditions under the action genre. It seems quite logical that when exposed to complex video materials that combine a range of aspects competing for working memory resources, learners would pay less attention to the aspect that they might perceive as insignificant. This perceived insignificance of the language aspect for learners with larger vocabulary sizes might stem from their ability to prioritize enjoyment in watching L2 videos. If they could comfortably understand the videos, they would probably prefer to engage with the story at a deeper level, rather than to look for unfamiliar words and

expressions, while learners with smaller vocabulary sizes might need all their cognitive resources focused on the language aspect in order to understand the content. This trade-off might be even more pronounced for such a dynamic genre as action, where, as we suggest, it was probably challenging to take in both action scenes and language elements, so participants might have been generally more inclined to choose one over the other.

In conclusion, our study has shown that sitcom and drama are more "neutral" genres, well suited for learning pragmatic formulaic sequences for various population groups. At the same time, the action genre has appeared as the most challenging one and possibly not providing optimal conditions for multimedia learning, and TED talks have proved to be more advantageous for older groups and learners with higher vocabulary knowledge. Although our study does not include exactly the same genres as Gilabert et al. (2018) and Suarez et al. (2021), it confirms that what they found for single words holds true for formulaic sequences: different genres provide different conditions for learning. The findings further overlap for the sitcom genre, which emerged as one of the most beneficial genres for incidental learning in the present study as well as in the previous ones.

4.4. Prior knowledge of the target items

Although not a part of the original research agenda, prior knowledge of the target formulaic sequences emerged as a strong predictor of learning gains, with participants who were more familiar with the items at the pre-test showing less improvement after the treatment. This seems to contradict Puimège and Peters (2020) and Majuddin et al. (2021), who found that learners could successfully build on their previous knowledge when exposed to meaningful audio-visual input. At first glance, this contradiction appears to be explained by either the

ceiling effect at the pre-test that did not leave much room for improvement, or the peculiarity of a scale-base test used to measure knowledge of the target items, where the psychological distance between the scale points increases up the scale. However, additional analysis ruled out the possibility of the ceiling effect, showing that even if the participants who scored close to the maximum were excluded, the negative effect of the pre-test scores remained significant. The scale-related explanation does seem to provide an accurate account of learners' progress on the vocabulary knowledge scale, since it is quite intuitive that participants moved easier from point 1 ("I have never seen this expression before") to point 2 ("I have seen this expression before, but I don't know what it means"; no translation required) than from point 2 to point 3 ("I have seen this expression before and I think I know what it means"; translation required). Surprisingly, although this explanation was valid in some cases, it was not consistently supported by the data either, since the pre- post-test gains for participants who reported no prior knowledge of some of the target items did not always stem from these items. Besides, this seemingly easy "jump" on the scale would have been balanced by the same amount of effort required to move from point 3 to point 4 ("I have seen this expression before, and I am sure what it means"). Therefore, another possible explanation is put forward to account for this finding. It seems that the more familiar learners were with the pragmatic formulaic sequences at the pre-test, the more confident they were about their knowledge, not realizing that this knowledge was sometimes incomplete or incorrect. This brings us back to the notion of deceptive transparency, used by Martinez and Murphy (2011) to explain learners' struggle with formulaic sequences consisting of familiar words. The pragmatic formulas used in the present study were comprised of high-frequent words that were most likely familiar to the participants, so they could certainly be regarded as deceptively transparent. The fact that learners with higher prior knowledge of the target

items showed less improvement after the treatment suggests that they were more susceptible to deceptive transparency, while learners with low scores were being more cautious in their interpretations. Additionally, since the last step on the knowledge scale was the ability to use the formulaic sequences sociopragmatically appropriately in a context, we can suggest that this probably was the most challenging aspect for learners. It could be the case that if they perceived the meaning of the items as known, whether correctly or not, low awareness of the pragmatic component and its importance possibly prevented them from adjusting their use based on the communicative situations they were exposed to (the "I know enough" attitude). Compared to Puimège and Peters' (2020) interpretation of the role of prior knowledge in learning of formulaic sequences, a slightly different picture emerged in the present study. The beneficial effect of repeated encounter with formulaic items in the meaningful input does not seem to accumulate in a linear manner, with each new encounter strengthening the previous knowledge, but rather appears to fluctuate as a wave depending on the amount of already existing knowledge and learners' awareness of what is still missing. Partial knowledge might be therefore a hindering factor in acquisition and not a facilitator unless learners are aware of its incompleteness. The difference between the described pattern and that suggested by Puimège and Peters (2020) could stem from the difference in the number of knowledge levels analyzed, since the scale used in the present study had 5 levels, and the authors of the previous study compared the two levels of meaning and form recall.

In sum, the negative effect of the pre-test scores in the present study might be attributed to the two following aspects of the learning process: the mental effort that was needed to deepen their knowledge was increasing with their progress on the selected measurement scale, and

participants' unawareness of the gap in their knowledge prevented them from building on the partial knowledge of the target items.

4.5. The role of metalinguistic awareness

Metalinguistic awareness was found to be a significant predictor of learning gains in our study, with participants who exhibited a higher level of this ability showing more improvement from the pre- to the post-test. Thus, our results support Kim's (2016) conclusion that language learners should be aware of the presence of various formulaic sequences in the input to be able to learn them incidentally. They are also in line with the pattern observed in Hubers et al. (2020), where L2 learners' intuitions about idiom properties were a reliable predictor of their idiom knowledge. It seems that the awareness of formulaic language and its main aspects helped participants to allocate their attention accordingly when watching the videos, so that they could pick up useful elements for further processing even after limited exposure. Therefore, we can conclude that it is important to be aware of what information can be extracted from the input for incidental learning to occur. This is a promising finding, since previous studies that looked into various awareness-raising activities were not able to establish conclusive links between learners' increased awareness and actual formulaic competence (for a review, see Boers & Lindstromberg, 2012). Although our design did not target unprompted production of formulaic sequences and rather focused on the initial stages of input processing involving intake, we can tentatively suggest that with sufficient exposure our finding could be extended to that aspect as well.

From the five components of metalinguistic awareness represented in the test, identifiability of formulaic sequences in the text appeared to have the strongest impact on learning gains,

confirming the importance of the noticing of formulaic items in the input in accordance with the Noticing Hypothesis (Schmidt, 1990). However, this component alone had lower explanatory power than in combination with the rest, which shows the validity of the multi-component approach to metalinguistic awareness.

It is also interesting to note that the relationship between metalinguistic awareness and proficiency, operationalized here as vocabulary size, was not quite straightforward. Although it was impossible to completely exclude purely linguistic aspects from the metalinguistic test that was designed for the purposes of the present study, the correlation between the two measures was very weak and got even weaker if we excluded the very low and very high vocabulary size scores. This shows that instead of being a simple by-product of proficiency, metalinguistic awareness is a varying individual ability, in line with early thoughts on the matter by Martinez and Murphy (2011) and Moskvina (2017).

Finally, it has to be acknowledged that the possibility of an unidentified confound affecting our results cannot be ruled out. Since the designed test followed a classical academic format, higher scores could be explained by general academic aptitude and level of engagement with the study as well. In other words, learners who were more used to taking academic tests, following complex instructions, and even using text redactors such as Microsoft Word might have had an advantage in taking the metalinguistic awareness test.

4.6. The role of statistical learning ability

Unlike metalinguistic awareness, learners' ability to pick up statistical regularities from the input did not contribute to learning of formulaic sequences from captioned videos. Several possible explanations can be proposed to interpret this finding. First of all, since the majority

of the results confirming the connection between statistical learning and language come from research on first language acquisition, our findings might suggest a different picture for second language learning. It seems that language learners in the instructional context do not tend to rely on implicit learning mechanisms. The majority of the target items were not completely unfamiliar to most of our participants, theoretically allowing them to build on previous encounters and to strengthen form-meaning links as well as intuitions about the boundaries of the formulaic sequences. Besides that, even those items that were reportedly new for the learners appeared four times during the study (pre-test, twice during the treatment, post-test). However, individual differences in tracking these encounters and absorbing the information that they provide did not effectively explain variation in learning outcomes, thus not corroborating the positive effect found at the level of processing in Kerz and Wiechmann (2019). It could be proposed that more consistent and extensive exposure is needed for language learners to profit from their statistical learning ability the way native speakers do. Therefore, a study involving a longitudinal design or an immersion context might further clarify whether statistical learning ability acquires more weight in the learning of formulaic sequences after a more prolonged exposure, or if it remains secondary to explicit learning mechanisms.

Finally, potential explanations stemming from our research design have to be acknowledged. Another design-related decision that might have affected our results is the choice of the test used to evaluate participants' ability to find regularities in the input. The Auditory-Verbal Adjacent (AVA) task from the artificial grammar learning paradigm was chosen because of its positive correlation with L2 processing of formulaic sequences from Kerz and Wiechmann (2019). Since it was the only study that had investigated statistical learning in the context of

L2 formulaic sequences, our research set out to elaborate on their findings. However, as mentioned in the literature review chapter, a later study with native English speakers by Isbillen et al. (2022) suggested a new measure of statistical learning ability specifically tailored for formulaic sequences, and their results showed that a new test was more strongly correlated with L1 processing of lexical bundles than the traditional AVA task. Perhaps by replicating the study for L2 learners we can further explore the statistical learning aspect in formulaic acquisition.

Concluding our comparison between the distributional and phraseological account of formulaic language we can suggest that despite the undoubted importance of distributional properties of formulaic items and the immense potential of statistical learning, L2 learners in the instructional context seem to rely more on explicit learning mechanisms involving metalinguistic awareness of formulaicity, and therefore such awareness should be promoted.

4.7. Components of metalinguistic awareness

Since the test of metalinguistic awareness designed for the purposes of the present study reflected four main aspects of formulaicity (identifiability, fixedness, meaning shifts, and compositionality) that were represented through five tasks, we were able to compare learners' performance across the formulaic components. The results suggest that fixedness, i.e., restricted or altogether absent variability in the form of formulaic sequences, was the easiest component for the participants, followed by compositionality and identifiability in the closed list, meaning shifts, and finally identifiability in the text. The fact that participants seemed to be more aware of fixedness than of other formulaic aspects is easily explained if we remember how this task was scored. All the answers that indicated learners' lack of

familiarity with the formulaic sequences used in the task were scored as missing values, because they were reflecting the absence of linguistic but not metalinguistic knowledge. Therefore, the scores used in the analysis come from the items that were sufficiently familiar to the learners. It can be inferred that awareness of the items' fixedness is closely connected to the knowledge of the items, and possibly forms an integral part of it. In other words, to know a formulaic sequence might also mean to be aware of the degree of its fixedness.

Compositionality and identifiability of formulaic sequences in the closed list of items were the next two easiest components of formulaicity, with no significant differences between them (section C in the results chapter). It suggests that these components involved a higher degree of abstraction and analysis of their previous language experiences on the part of the learners than fixedness.

In line with Guz (2014), who reported low scores from advanced language learners on a similar task, identifiability of formulaic sequences in a text was the most challenging component for our participants, closely followed by meaning shifts. Learners' poor performance on the task of detecting formulaic items in a text also confirms Bishop's (2004) findings that language learners generally struggle with identifying formulaic sequences as special units in the input. It is not surprising that participants found meaning shifts to be a difficult aspect of formulaicity as well, since it required a breakdown of already familiar collocations into constituents for further analysis, which assumes a quite high level of abstraction.

4.8. Metalinguistic awareness across formulaic types

To examine learners' idea of formulaicity, their metalinguistic awareness was compared across the formulaic types for each of the components. The formulaic types included collocations, idioms, binomials, pragmatic formulaic sequences, lexical bundles, and phrasal verbs. It is interesting to note that the two tasks devoted to identifiability, involving a text and a closed list correspondingly, exhibited different patterns. In the text, idioms and phrasal verbs were identified most consistently, followed by collocations; while lexical bundles, binomials, and pragmatic formulaic sequences were the most commonly overlooked types. In the list, binomials remained mostly undetected and phrasal verbs held their position as a highly identifiable type, while idioms were identified considerably less often than in the text, and the chances of the other items (lexical bundles, pragmatic formulaic sequences, collocations) to be detected got higher. These changes convey the general idea that identifiability of formulaic sequences is higher in a finite list of items, since learners do not have to segment a text into smaller units first. Additionally, it indicates that phrasal verbs are identified as formulaic due to the peculiarity of their form rather than meaning, since they are equally successfully detected in the context and without it, while the opposite is the case for idioms, which were less identifiable in the absence of context which would have highlighted their non-compositional meanings. Finally, it seems that binomials tend not to be recognized by language learners as a special class.

Learners' awareness of the fixedness aspect was at a comparable level for collocations, idioms, and pragmatic formulaic sequences, but was significantly lower for binomials. It might indicate that participants did not consider binomials to be a part of the formulaic continuum and were therefore unaware of the importance of word order for this formulaic

type. In terms of collocations, our results do not seem to be in line with Fioravanti et al. (2021), where L2 learners exhibited lower sensitivity to lexical fixedness of collocations compared to that of idioms. This is possibly explained by the fact that their study applied a more flexible measure of scale, while in the present dissertation a more rigid task based on multiple choice was used.

As for compositionality, language learners' sensitivity to this formulaic aspect did not differ much between collocations, phrasal verbs, and pragmatic formulaic sequences, however, it was significantly higher for idioms, confirming that this type was recognized as prototypically formulaic by participants.

Overall, these results indicate that participants' idea of what comprises formulaic language is formed around the traditional core of idiomatic expressions, and less figurative types, especially binomials, lexical bundles, and pragmatic formulaic sequences, are largely left outside of learners' formulaic scope. This corroborates the survey findings from Kim (2016), which showed that learners' idea of formulaic sequences was limited to figurative idioms and proverbs.

4.9. Limitations and further research

Although we are excited to share our results and believe that they make a significant contribution to our understanding of formulaic language and its acquisition, the limitations of the present study have to be acknowledged.

First of all, due to the post-COVID restrictions our data was collected online. An effort was made to ensure that the instructions were followed and the used instruments allowed a certain degree of control over the procedure (learners could not skip the videos, test logs were

available, etc.), however, we cannot state with the absolute certainty that participants' attention was undivided or that the experimental conditions across learners were exactly the same. A replication of the study in a more controlled lab or classroom environment could help to validate our claims.

Second, since we had 4 experimental groups, the number of participants per group was quite low (26-28). It is possible that certain effects did not reach a significant level due to a low number of observations, and our findings need to be confirmed on a larger scale. It is also noteworthy that our research involved only 4 main genres. While it offered a good starting point for the investigation of learning conditions under different genres, a larger set of genres should be examined, as well as the subcategories within a genre.

A number of limitations arise from our intent to create highly controlled experimental conditions (at least at the level of materials). As we wanted to compare learning of the same formulaic sequences across the genres, with the same number of occurrences, and keep the videos representative of the selected genres and accessible for learners, the number of target items was small (6 formulaic sequences), and the overall exposure limited (approximately 30 minutes in total, each item appeared twice). For the same reasons our treatment was a video sequence composed of individual clips rather than full episodes of TV programs, which might have created slightly different conditions. Another research design, possibly longitudinal, could be advised to explore more freely which formulaic sequences occur naturally in the TV programs of different genres and how they can be learnt from the exposure to the L2 videos of these genres. This would also allow us to look at other factors, such as number of occurrences, formulaic type, and distributional properties (phrasal frequency, mutual information). The present study focused exclusively on one type of formulaic sequences,

namely pragmatic formulas, and consistent comparison between different formulaic types with a balanced number of occurrences could provide interesting insights into the general potential of L2 videos for formulaic acquisition and genre-related differences in that process.

The pre-test scores in our study suggested that prior knowledge of the selected target items varied among participants. It allowed us to evaluate the effect of this variation on learning, however, having only completely unknown or only already partially known items might have led to stronger claims about what type of learning captioned videos can and cannot promote. In general, learning was operationalized here as intake, and the initial stages of input processing were of our particular interest. As a result, we did not administer a delayed post-test and the productive knowledge of formulaic sequences was not deeply explored. Further research specifically focusing on the effect of L2 videos on the production of formulaic sequences is suggested.

Additionally, participants' proficiency was estimated through their vocabulary size. Perhaps, a more comprehensive proficiency measure (e.g., Oxford Placement Test) could have described proficiency variation within the sample more accurately. Overall, the sample in the present study was characterized by a high heterogeneity in terms of age and vocabulary size. While it accurately reflected a real classroom demographic of public language schools in Catalonia and was advantageous for some aspects of our analysis, it definitely limits generalizability of our research. Further research separately exploring particular age and proficiency groups could be advised. Last but not least, our participants were volunteers, which might suggest a generally higher level of engagement, and the results should be confirmed for intact classes and random samples.

Regarding individual differences addressed in the study, as mentioned in the discussion chapter, statistical learning ability should be explored more to identify the best suited measure to capture the individual differences in this ability. Although the choice of the test applied in our study was based on the previous findings, a combination of different measures could have yielded stronger results. We also suggest extending the research agenda to other cognitive individual differences, such as working memory or attention. Besides, since our data was collected online, it relies to a large extent on participants' self-discipline and diligence. For a sensitive, cognitively demanding measure such as statistical learning ability, that might have resulted in some additional data noise, and more controlled lab conditions could have been preferable.

Finally, the multicomponent measure of metalinguistic awareness was proposed and applied here for the first time. More refinement and further exploration of this operationalization of the studied concept is needed to make our conclusions more robust.

4.10. Conclusions and pedagogical implications

This doctoral dissertation contributes to our understanding of the acquisition of formulaic sequences from audio-visual input and brings into focus earlier unexplored factors of genre, metalinguistic awareness, and statistical learning ability.

Captioned videos were shown as a valid tool for promoting learning of pragmatic formulaic sequences, confirming the results previously obtained for other formulaic types, which were mostly collocations, phrasal verbs and idioms. It suggests that captioned videos in the target language can be used as a part of multimedia instruction for such sequences, whether during classroom viewing or as an extracurricular activity.

However, both instructors and learners may need to consider genre when selecting L2 video materials, since our data suggests that different genres might provide different learning conditions and vary in their suitability for certain population groups as well. According to our findings, sitcom and drama appear to be well suited for the general population, regardless of their age and vocabulary knowledge. Since the sitcom group enjoyed their viewing experience the most and the drama one the least, sitcom could be recommended as the most advantageous genre for multimedia instruction. Regarding the other two genres, TED talks seem to be beneficial for older learners or those with a higher vocabulary level, while action is more appropriate for younger learners or learners with less vocabulary knowledge. Overall, action was demonstrated in this thesis to be the least beneficial genre for learning, therefore it probably should be used cautiously in multimedia instruction for formulaic sequences or avoided altogether. Additionally, if considering using any genre not covered in the research, its potential for learning could be evaluated though the conditions specified in the present study.

Our findings further suggest that L2 learners might tend to rely on more explicit learning mechanisms in the process of incidental learning of formulaic sequences from audio-visual input, or at least that they do not seem to draw on their implicit statistical learning ability in that process. Metalinguistic awareness was a significant predictor of learning gains, while no effect of statistical leaning ability was observed. Therefore, learners' awareness of various sequences that comprise formulaic language and their main features, such as fixedness, compositionality, and meaning shifts, is important for successful acquisition of such sequences. Learners need to know what language elements are present in the input and which aspects of these elements are most relevant in order to allocate their attention efficiently

during input processing. In the case of formulaic sequences, they need to know that the language they are exposed to contains not only individual words but also longer word combinations, and that these combinations are not formed randomly at the moment of a conversation but rather reoccur across conversations. Besides, they typically reoccur in the exact the same or a very similar form, and their meaning organization is different from other word groups, where the elements are combined freely. To conclude, our findings emphasize the significance of awareness in formulaic acquisition and that of awareness-raising activities that might enhance it. Even if these activities do not result in immediate improvement within a language course, as it was the case, for example, in Jones and Haywood (2004), the increased awareness they promote might provide a steady basis for further L2 development. Combined with a powerful source of authentic input, such as immersion in the target language community or, as it was demonstrated here, captioned TV series and programs, this awareness can lead to increased uptake of formulaic sequences. Therefore, audio-visual input and its widespread accessibility might address the issue of limited exposure raised by Boers and Lindstromberg (2012).

Furthermore, it was demonstrated here that partial knowledge of pragmatic formulaic sequences could have a hindering effect on further improvement of that knowledge. Therefore, learners' attention should be drawn to the final objective of acquisition of such sequences, i.e., correct sociopragmatic use. It is also advisable to inform learners through explicit instruction that despite the fact that pragmatic formulas often consist of high frequent, familiar words, they are a special class of items, and their interpretation and use should follow certain rules, different from those for free combinations.

Finally, a multicomponent operationalization of metalinguistic awareness of formulaicity was proposed and implemented in this doctoral dissertation for the first time. It was shown that metalinguistic awareness is not a simple by-product of proficiency and can be examined at the level of individual differences. In addition, based on the collected data, learners' idea of formulaicity appears to be somewhat limited and centered around the most figurative items, such as idioms, while lexical bundles, pragmatic formulaic sequences, and especially binomials are not consistently recognized and treated as formulaic. It could be therefore proposed that awareness-raising instruction should focus on expanding learners' idea of what items comprise formulaic language and what is special about each of the item types.

Similarly, learners' awareness of the essential component of formulaicity identified in the present study (fixedness, compositionality, identifiability, and meaning shifts) was not consistent. They struggled most with identifying formulaic sequences in a text, and since this component had the biggest impact on learning from the audio-visual treatment, it confirms the validity of awareness-raising activities aimed at this formulaic aspect, such as text chunking. Another possible tool for directing learners' attention to formulaic sequences in the input is typographic enhancement, which proved to be valid for this purpose in a number of previous studies (Bishop, 2004; in the audio-visual input: Moskvina, 2017; Majuddin et al., 2021; Puimège et al., 2021). Besides, fixedness was shown to have a close connection with the knowledge of formulaic items, suggesting that this aspect might need to be emphasized in the instruction as well.

In sum, this doctoral dissertation provides new insights into the acquisition of formulaic language from audio-visual input and the ways it could be enhanced through the careful choice of video materials as well as through learner-focused instruction. L2 learners' idea of

formulaicity was explored and appeared to be incomplete, with further recommendation on possible ways to broaden it provided.

References

Alderson, J. C., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language teaching research*, *1*(2), 93-121. https://doi.org/10.1177/136216889700100202

Allerton, D.J., Nesselhauf, N., & Skandera, P. (Eds.). (2004). *Phraseological Units: Basic Concepts and their Application*. Schwabe.

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Beninca, P., & Semenza, C. (2012). Is "hit and run" a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology*, *3*(11), 1–11. https://doi.org/10.3389/fpsyg.2012.00011

Bardovi-Harlig K. (2008). Recognition and production of formulas in L2 pragmatics. In Han Z.-H. (Ed.), *Understanding second language process* (pp. 205–222). Clevedon, UK: Multilingual Matters.

Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning*, *59*(4), 755-795. https://doi.org/10.1111/j.1467-9922.2009.00525.x

Bardovi-Harlig, K. (2012). Formulas, routines, and conventional expressions in pragmatics research. *Annual Review of Applied Linguistics*, *32*, 206-227. https://doi.org/10.1017/S0267190512000086

Bardovi-Harlig, K., & Vellenga, H. E. (2012). The effect of instruction on conventional expressions in L2 pragmatics. *System*, *40*(1), 77-89. https://doi.org/10.1016/j.system.2012.01.004

Barón J, & Celaya, M.L. (2022). 'May I do something for you?': The effects of audio-visual material (captioned and non-captioned) on EFL pragmatic learning. *Language Teaching Research*, *26*(2), 238–255. https://doi.org/10.1177/13621688211067000

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, *26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International journal of corpus linguistics*, *14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, *25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In Schmitt, N. (Ed.), *Formulaic sequences* (pp. 19-37). John Benjamins.

Bogaerts, L., Siegelman, N., Christiansen, M. H., & Frost, R. (2022). Is there such a thing as a "good statistical learner"? *Trends in cognitive sciences*, *26*(1), 25-37. https://doi.org/10.1016/j.tics.2021.10.012

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the

test. *Language teaching research*, *10*(3), 245-261. https://doi.org/10.1191/1362168806lr195oa

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83-110. https://doi.org/10.1017/S0267190512000050

Brice, H., Siegelman, N., Van den Bunt, M., Frost, S., Rueckl, J., Pugh, K., & Frost, R. (2022). Individual differences in L2 literacy acquisition: predicting reading skill from sensitivity to regularities between orthography, phonology, and semantics. *Studies in Second Language Acquisition*, *44*(3), 737-758. https://doi.org/10.1017/S0272263121000528

British council (n.d.). *A2 reading: an email from a friend*. https://learnenglish.britishcouncil.org/skills/reading/a2-reading/email-friend

Burger, H., Dobrovol'skij, D., Kühn, P., & Norrick., N.R. (Eds.). (2007). *Phraseology: an International Handbook of Contemporary Research*. Mouton de Gruyter.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95-122. https://doi.org/10.1177/0023830918823230

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(3), e1460.

Cintrón-Valentín, M., García-Amaya, L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal*, *47*(4), 1–21. https://doi.org/10.1080/09571736.2019.1615978

Columbus, G. (2010). Processing MWUs: Are MWU subtypes psycholinguistically real. In Wood, D. (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 194-210). Bloomsbury Publishing.

Columbus, G. (2013). In support of multiword unit classifications: Corpus and human rating data validate phraseological classifications of three different multiword unit types. *Yearbook of Phraseology*, *4*(1), 23-44. https://doi.org/10.1515/phras-2013-0003

Cook, R. D. (2011). Cook's distance. In *International Encyclopedia of Statistical Science* (pp. 301-302). Springer.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*(5), 570-590. https://doi.org/10.1093/applin/amt056

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second language research*, *26*(2), 163-188. https://doi.org/10.1177/0267658309349431

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceeding of Fourth International Conference on Spoken Language Processing*, 3, 1393-1396. https://doi.org/10.1109/ICSLP.1996.607874

Elder, C., & Manwaring, D. (2004). The relationship between metalinguistic knowledge and learning outcomes among undergraduate students of Chinese. *Language Awareness*, *13*(3), 145-162. https://doi.org/10.1080/09658410408667092

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in second language acquisition*, *18*(1), 91-126. https://doi.org/10.1017/S0272263100014698

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375–396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, *20*(1), 29-62. https://doi.org/10.1515/text.1.2000.20.1.29

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2020). Vocabulary Learning Through Viewing Captioned or Subtitled Videos and the Role of Learner- and Word-Related Factors. *CALICO Journal*, *37*(3), 233–253. https://doi.org/10.1558/cj.39370

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2023). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning*, *36*(1-2), 26-51. https://doi.org/10.1080/09588221.2021.1899244

Fioravanti, I., Senaldi, M. S. G., Lenci, A., & Siyanova-Chanturia, A. (2021). Lexical fixedness and compositionality in L1 speakers' and L2 learners' intuitions about word

combinations: Evidence from Italian. *Second Language Research*, *37*(2), 291-322. https://doi.org/10.1177/0267658320941560

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological science*, *24*(7), 1243-1252. https://doi.org/10.1177/0956797612472207

Frumuselu, A. D., De Maeyerb, S., Doncheb, V., Gutiérrez-Colon Plana, M. Del M. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, *32*(B), 107–117. https://doi.org/10.1016/j.linged.2015.10.00

Geeraerts, D. (1995). Specialization and reinterpretation in idioms. In M. Everaert, E. J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 57–73). Erlbaum.

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29-44. https://doi.org/10.1016/j.system.2016.04.004

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign language annals*, *24*(3), 239-258. https://doi.org/10.1111/j.1944-9720.1991.tb00469.x

Gass, S. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates Publishers.

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning &Technology*, *23*(2), 84–104.

Gesa, F., & Miralpeix, I. (2022). Effects of watching subtitled TV series on foreign language vocabulary learning: does learners' proficiency level matter? In C. Lütge (Ed.), *Foreign Language Learning in the Digital Age. Theory and Pedagogy for Developing Literacies* (pp. 159–173). Routledge. https://doi.org/10.4324/9781003032083-14

Gibbs Jr, R. W. (1991). Semantic analyzability in children's understanding of idioms. *Journal of Speech, Language, and Hearing Research*, *34*(3), 613-620. https://doi.org/10.1044/jshr.3403.613

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, *87*(6), 1082-1116.

Gilabert Guerrero, R., Suárez, M. D. M., Moskvina, N., Levkina, M., Barón Parés, J., Vasylets, O., & Feijóo Antolín, S. (2018, September 5-8). *The effects of captioned video on processing and vocabulary learning under different genres* [Conference presentation]. The 28th Conference of the European Second Language Association (EuroSLA), Münster, Germany.

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL-international journal of applied linguistics*, *166*(1), 94-126. https://doi.org/10.1075/itl.166.1.03fer

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27–49). John Benjamins.

Heatley, A., Nation, P. & Coxhead, A. 2002. RANGE [Computer software]. Available at: http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

Guz, E. (2013). Gauging advanced learners' language awareness: Some remarks on the perceptual salience of formulaic sequences. In Łyda, A., & Szcześniak, K. (Eds.), *Awareness in Action: The Role of Consciousness in Language Acquisition* (pp. 165-181). Springer International Publishing.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*(2), 296-323. https://doi.org/10.1111/lang.12143

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. https://doi.org/10.1080/23273798.2016.1152389

Herrarte, D. L. (1998). Metalinguistic awareness and the learning of English as an L3. *Atlantis*, 69-79.

Hubers, F. (2019). *Two of a kind: Idiomatic expressions by native speakers and second language learners* [Doctoral dissertation, Radboud University]. Netherlands Graduate School of Linguistics. https://www.lotpublications.nl/two-of-a-kind-idiomatic-expressions-by-native-speakers-and-second-language-learners

Hubers, F., Cucchiarini, C., & Strik, H. (2020). Second language learner intuitions of idiom properties: What do they tell us about L2 idiom knowledge and acquisition? *Lingua*, *246*, 102940. https://doi.org/10.1016/j.lingua.2020.102940

IMDB. (n.d.). *The African Queen*. https://www.imdb.com/title/tt0043265/

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition, 225,* 105123. https://doi.org/10.1016/j.cognition.2022.105123

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In Schmitt, N. (Ed.), *Formulaic sequences* (pp. 269-300). John Benjamins.

Kecskes, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of pragmatics*, *32*(5), 605-625. https://doi.org/10.1016/S0378-2166(99)00063-6

Kecskes, I. (2007). Formulaic language in English lingua franca. *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects*, *1*, 191-218. https://doi.org/10.1515/9783110198843

Kecskes, I. (2010). Situation-bound utterances as pragmatic acts. *Journal of pragmatics*, *42*(11), 2889-2897. https://doi.org/10.1016/j.pragma.2010.06.008

Kecskes, I., Obdalova, O., Minakova, L., & Soboleva, A. (2018). A study of the perception of situation-bound utterances as culture-specific pragmatic units by Russian learners of English. *System, 76*, 219-232. https://doi.org/10.1016/j.system.2018.06.002

Kerz, E., Wiechmann, D. (2019). Effects of Statistical Learning Ability on the Second Language Processing of Multiword Sequences. In: Corpas Pastor, G., Mitkov, R. (Eds.), *EUROPHRAS: International Conference on Computational and Corpus-Based Phraseology* (pp. 200-214). Springer. https://doi.org/10.1007/978-3-030-30135-4_15

Khazdouzian, Y., Celaya, M.L., Barón, J. (2021). When watching is not enough: The effects of captions on L2 pragmatics acquisition and awareness. *RAEL - Revista Electrónica de Lingüística Aplicada, 19,* 90–107.

Kim, C. (2016). L2 learners' recognition of unfamiliar idioms composed of familiar words. *Language Awareness*, *25*(1-2), 89-109. https://doi.org/10.1080/09658416.2015.1122025

Kock, N., & Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for information Systems*, *13*(7), 546-580. https://doi.org/10.17705/1jais.00302

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research & Development*, *47*(1), 51-60.

Laufer, B. (1989a). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, *6*, 10-20.

Lee, M. & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: an eye-tracking study. *The Modern Language Journal, 102*(3), 557–577. https://doi.org/10.1111/modl.12503

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language learning*, *54*(2), 193-226. https://doi.org/10.1111/j.1467-9922.2004.00254.x

Lin, P. M. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, *42*, 164-176. https://doi.org/10.1016/j.system.2013.11.010

Lin, P. (2018). Formulaic language and speech prosody. In Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 78-94). Routledge.

Madarbakus-Ring, N. (2020). Developing graded TED Talks to integrate academic vocabulary into listening lessons for pre-sessional learners. *The TESOL encyclopedia of English language teaching*, 1-7.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition, 43*(5), 985-1008. https://doi.org/10.1017/S0272263121000036

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *Tesol Quarterly*, *45*(2), 267-290. https://doi.org/10.5054/tq.2011.247708

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*, 314-324.

Matielo, R., Pires de Oliveira, R., & Baretta, L. (2018). Subtitling, worling memory, and L2 learning: A correlational study. *Revista Brasileira de Linguística Aplicada, 18*(3), 665–696. https://doi.org/10.1590/1984-6398201812773

Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., ... & Weinberger, D. R. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience letters*, *392*(1-2), 32-37. https://doi.org/10.1016/j.neulet.2005.09.025

Mayer, R. E. (2009). *Multimedia learning (2nd ed.).* New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R.E., Fiorella, L. & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development, 68*, 837–852. https://doi.org/10.1007/s11423-020-09749-6

Meara, P. (2022, September 2). *Free software from _lognostics*. _Lognostics. https://www.lognostics.co.uk/tools/index.htm

Meara, P. and Jones, G. (1990). Eurocentres Vocabulary Size Test, 10KA (Zurich: Eurocentres), *TESL Canadian Journal, 3*(1), pp. 69-79.

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary* (Vol. 105). Multilingual Matters.

McDonough, K., & Trofimovich, P. (2016). The role of statistical learning and working memory in L2 speakers' pattern learning. *The Modern Language Journal*, *100*(2), 428-445. https://doi.org/10.1111/modl.12331

Miller, G. (1956). Human memory and the storage of information. *IRE Transactions on Information Theory*, *2*(3), 129-137.

Mollin, S. (2012). Revisiting binomial order in English: ordering constraints and reversibility. *English Language & Linguistics*, *16*(1), 81-103. https://doi.org/10.1017/S1360674311000293

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749-773. https://doi.org/10.1017/S0272263119000706

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118-141.

Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In Cowie, A.P. (Ed.), *Phraseology. Theory, Analysis, and Applications*, 79–100. Oxford University Press.

Moskvina, N. (2017). *Noticing and intake of formulaic sequences in L2 subtitles: an eye-tracking study on the effects of genre and input enhancement* [Unpublished master's thesis]. University of Barcelona.

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, *67*, 77-86. https://doi.org/10.1016/j.system.2017.04.015

Muñoz, C., Pattemore, A., & Avello, D. (2022). Exploring repeated captioning viewing as a way to promote vocabulary learning: time lag between repetitions and learner

factors. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2113898

Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Available from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. John Benjamins. https://doi.org/10.1075/scl.14

Netflix. (n.d.). *Timed Text Style Guides*. Retrieved on December 12, 2020, from https://partnerhelp.netflixstudios.com/hc/en-us/categories/1500000000781-Timed-Text-Resources

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading research quarterly*, 95-106. https://doi.org/10.2307/747835

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298-320. https://doi.org/10.1177/1362168816639619

Nguyen, C. D., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED talk. *Tesol Quarterly*, *53*(1), 5-29. https://doi.org/10.1002/tesq.441

Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, *40*(5), 816-833. https://doi.org/10.1093/applin/amy027

Nurmukhamedov, U. (2017). Lexical coverage of TED Talks: Implications for vocabulary instruction. *TESOL Journal*, *8*(4), 768-790. https://doi.org/10.1002/tesj.323

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *Quarterly Journal of Experimental Psychology*, *69*(12), 2390-2401. https://doi.org/10.1080/17470218.2015.1112825

Pattemore, A., & Muñoz, C. (2020). Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. *System, 93*, 102303.

Pattemore, A., Suárez, M. D. M., & Muñoz, C. (2021, June 30-July 3*). Learning L2 grammar constructions through audio-visual input: The role of aptitude* [Conference presentation]. The 30th Conference of the European Second Language Association (EuroSLA 30), Barcelona, Spain.

Pattemore, A., & Muñoz, C. (2022). Captions and Learnability Factors in Learning Grammar from Audio-Visual Input. *JALT CALL Journal*, *18*(1), 83-109.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32,* 130-149. https://doi.org/10.1017/S0267190512000098

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–225). Longman

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113-138. https://doi.org/10.1177/1362168814568131

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134-148.

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *Tesol Quarterly*, *53*(4), 1008-1032. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551-577. https://doi.org/10.1017/S0272263117000407

Pollio, H.., Barlow, J., Fine, H.., & Pollio, M. (1977). *Psychology and the Poetics of Growth: Figurative Language in Psychology, Psychotherapy, and Education*. Lawrence Erlbaum Associates.

Puimège, E., & Peters, E. (2019a). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424-438. https://doi.org/10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*(3), 525-549. https://doi.org/10.1017/S027226311900055X

Puimège, E., Montero Perez, M., & Peters, E. (2023). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, *39*(2), 471-492. https://doi.org/10.1177/02676583211049741

Pujadas, G. & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, *47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, *42*(3), 551–575. https://doi.org/10.1017/S0272263120000042

Reinders, H. W. (2012). Towards a definition of intake in second language acquisition. *Applied Research on English Language*, *1*(2), 15-36.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical 190 coverage, attitudes, and captions* [Doctoral thesis, Victoria University of Wellington]. Digital repository. http://hdl.handle.net/10063/2870

Rodgers, M. P. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL-International Journal of Applied Linguistics*, *169*(1), 191-211. https://doi.org/10.1075/itl.00012.rod

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics*, *29*(2), 173-199. https://doi.org/10.1093/applin/amm037

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics*, *11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid. In Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127-151). John Benjamins.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of memory and language*, *81*, 105-120. https://doi.org/10.1016/j.jml.2015.02.001

Siegelman, N., Rueckl, J. G., Steacy, L. M., Frost, S. J., van den Bunt, M., Zevin, J. D., Seidenberg, M. S., Pugh, K. R., Compton, D. L., & Morris, R. D. (2020). Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *Journal of Memory and Language, 114*, 104145. https://doi.org/10.1016/j.jml.2020.104145

Sinclair, J. M. (1991). *Corpus, concordance, collocation*, Oxford: Oxford University Press.

Sinclair, J. M. (2007a). 'Collocation reviewed' (manuscript), Tuscan Word Centre, Italy.

Sinclair, J. M. (2007b). 'Defining the definiendom—new' (manuscript), Tuscan Word Centre, Italy.

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (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. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549-569. https://doi.org/10.1093/applin/amt054

Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.). (2018). *Understanding formulaic language: A second language acquisition perspective*. Routledge. https://doi.org/10.4324/9781315206615

Sorhus, H. B. (1977). To Hear Ourselves--Implications for Teaching English as a Second Language. *English Language Teaching Journal*, *31*(3), 211-221.

Suárez, M. D. M., Gilabert, R., & Moskvina, N. (2021). The mediating role of vocabulary size, working memory, attention and inhibition in early vocabulary learning under different TV genres: An exploratory study. *TESOL Journal*, *12*(4), 1– 29. https://doi.org/10.1002/tesj.637

Supasiraprapa, S. (2019). Frequency effects on first and second language compositional phrase comprehension and production. *Applied Psycholinguistics*, *40*(4), 987-1017. https://doi.org/10.1017/S0142716419000109

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73.

Thewissen, J. (2008). The phraseological errors of French-, German-, and Spanish speaking EFL learners: Evidence from an error-tagged learner corpus. In *Proceedings from the 8th Teaching and Language Corpora Conference (TaLC8), Portugal*, 300–306.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. R. Heredia & A. B. Cieślicka (Eds.),

Bilingual Figurative Language Processing (pp. 171–207). Cambridge University Press. https://doi.org/10.1017/CBO9781139342100.011

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT journal*, *42*(4), 272-281. https://doi.org/10.1093/elt/42.4.272

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning.* Palgrave Macmillan. https://doi.org/10.1057/978-1-137-50045-8

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied linguistics*, *34*(4), 457-479. https://doi.org/10.1093/applin/ams074

Vetchinnikova, S. (2014). *Second language lexis and the idiom principle* [Doctoral thesis, University of Helsinki]. Digital repository. http://hdl.handle.net/10138/135691

VideoLan. (2006). *VLC media player.* Retrieved from https://www.videolan.org/vlc/index.html

Vu, D. V., Noreillie, A. S., & Peters, E. (2023). Incidental collocation learning from reading-while-listening and captioned TV viewing and predictors of learning gains. *Language Teaching Research*, 13621688221151048. https://doi.org/10.1177/13621688221151048

Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. Language Learning. https://doi.org/10.1111/lang.12495

Wang, X., Boers, F., & Warren, P. (2022). Prompting language learners to guess the meaning of idioms: do wrong guesses linger? *Language Awareness*, 1-16. https://doi.org/10.1080/09658416.2022.2153859

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13-40. https://doi.org/10.3138/cmlr.53.1.13

Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies, 11*(1), 117–136.

Webb, S., & Rodgers, M. P. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407-427. https://doi.org/10.1093/applin/amp010

Webb, S., & Rodgers, M. P. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335-366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition, 42*(3), 599–624. http://doi.org/10.1017/S0272263120000029

Woll, N. (2018). Investigating dimensions of metalinguistic awareness: what think-aloud protocols revealed about the cognitive processes involved in positive transfer from L2 to L3. *Language Awareness*, *27*(1-2), 167-185.

https://doi.org/10.1080/09658416.2018.1432057

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual review of applied linguistics, 32*, 231-254. https://doi.org/10.1017/S026719051200013X

Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. In Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19-37). Routledge.

Appendices

Appendix A. Vocabulary knowledge scale (pre- and post-test)

1. Be my guest

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don't know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

2. Let's face it.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don't know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

3. Good for you.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

4. Give me a break

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

5. Have a nice day.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

6. You don’t say.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

7. Let it go.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

8. What's the big deal?

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don't know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

9. Take care.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don't know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

10. Never mind.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

11. Watch out!

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

12. Talk to you soon.

☐ I have never seen this expression before (I don't remember).

☐ I have seen this expression before, but I don’t know what it means.

☐ I have seen this expression before, and I think it means:
Please, write a synonym or translation/explanation

.....

☐ I know this expression. It means:
Please, write a synonym or translation/explanation

.....

☐ I can use this expression in a context:
(Please, write a short dialogue: 2-3 sentences)

(If you do this section, please also do Section 4)

.....

.....

.....

Appendix B. TV programs and episodes used in the study.

Sitcom:

- “Joey”. Season 1 episodes 2 (Joey and the Student), 16 (Joey and the Tonight Show), 17 (Joey and the Valentine’s Date); season 2 episode 3 (Joey and the Spanking).
- “Friends”. Season 1 episode 17 (The One with Two Parts: Part 2); season 2 episode 10 (The One with Russ); season 3 episode 4 (The One with the Metaphorical Tunnel); season 4 episode 5 (The One with Joey's New Girlfriend); season 8 episode 8 (The One with the Stripper); season 9 episodes 8 (The One with Rachel's Other Sister) and 15 (The One with the Mugging).

Drama:

- “Gossip girl”. Season 3 episodes 12 (The Debarted) and 18 (The Unblairable Lightness of Being); season 4 episode 4 (Touch of Eva).
- “90210”. Season 1 episodes 3 (Lucky Strike), 4 (Wide Awake and Dreaming), 11 (Hello, Goodbye, Amen), 17 (Off the Rails); season 2 episodes 8 (Women's Intuition), 20 (Meet the Parent); season 5 episode 21 (Scandal Royale).

TED talks:

- America Ferrera: My identity is a superpower - not an obstacle.
- Rick Warren: A life of purpose.
- Noah Wilson-Rich: How you can help save the bees, one hive at a time.
- Russ Altman: What really happens when you mix medications?
- Billy Collins: Everyday moments, caught in time.
- Sandra Aamodt: Why dieting doesn't usually work.

- Sheena Iyengar: The art of choosing.
- Margaret Heffernan: The dangers of willful blindness.
- Molly Crockett: Beware neuro-bunk
- Julie Lythcott-Haims: How to raise successful kids -- without over-parenting
- Susan David: The gift and power of emotional courage
- Shonda Rhimes: My year of saying yes to everything

Action:

- The Flash: Season 3 episode 7 (Killer Frost); season 5 episodes 10 (The Flash & The Furious), 19 (Who Is Harrison Wells); season 6 episodes 1 (Into the Void), 18 (Pay the Piper).
- Arrow: season 2 episodes 1, 19, 23; season 4 episode 18 (Eleven-Fifty-Nine); season 5 episode 8 (Invasion!).
- Daredevil (the movie).
- Legends of Tomorrow (Pilot).

Appendix C. Links to the experimental treatment

Sitcom group: https://edpuzzle.com/assignments/618034a2b400bd415e81841e/watch

TED talks group: https://edpuzzle.com/assignments/618034c030c765416b59d7af/watch

Drama group: https://edpuzzle.com/assignments/618034d75f975a419ae9dfe2/watch

Action group: https://edpuzzle.com/assignments/618034eb73369041482fd834/watch

Appendix D. Multicomponent test of metalinguistic awareness

Task1. *Underline groups of two or more words that you think are a* ***fixed unit*** *(and not just individual words). Don't worry if you are not quite familiar with the concept of "fixed units". We are interested in how* ***you*** *would interpret it.*

Hi Samia,

How are you doing? Just a quick email to say that meeting this weekend sounds like a great idea. Saturday is better for me because I'm meeting my parents on Sunday. So, if that's still good for you, why don't you come here? Then you can have a look at the new flat and all the work we've done on the kitchen since we moved in. It will blow your mind! Tom has been working on it day and night for two months. We can eat at home and then go to the park in the evening. I think we will eat around 8. Come earlier if you want to give me a hand with dinner. We can have some wine! It's going to be so good to catch up finally. I want to hear all about your new job! I didn't get a chance to congratulate you yet. You totally deserve it after all the hard work. By the way, I invited Kevin too, but he is not sure if he is going to come. He has been a real couch potato lately. He says he needs some peace and quiet now.

Our address is 52 Charles Road, but it's a bit difficult to find because the house numbers are really strange here. If you turn left at the post office and keep going past the big white house on Charles Road, there's a small side street behind it with the houses 50–56. Don't ask me why the side street doesn't have a different name! Mary came last week and she was walking up and down the street for half an hour before she found it. So call me if you can't find us and I'll come and get you. I'll get back to you about the menu later. I want to make tacos or something like that. Do you have any allergies?

Really looking forward to seeing you!

See you soon!

Kate

Task 2. *Underline **all** the expresions from the list that are **fixed.** Only choose the ones that you sure about.*

1. Forget about
2. I don't eat meat.
3. It's raining cats and dogs.
4. Don't burn the pie!
5. Pick up
6. Have apples
7. You're welcome.
8. She is tall and pretty.
9. Right and wrong
10. For my heart
11. Are you kidding me?
12. He wrote her a letter.
13. Lost and found
14. Salty and tasty
15. As well as
16. Make sense
17. Cool as a cucumber
18. To begin with
19. Stop for
20. Banana and honey
21. Shut up
22. Candy recipe
23. Coffee break
24. To that tree

Task 3. *Take a look at the words in **bold**. Choose the two sentences where the marked words have the same meaning. Take into account the context that they appear in! If you don't know the meaning of some of the phrases, choose the option "I don't know".*

For example:

a. *Monica is a very **beautiful** woman.* ☒
b. *I am reading a very **interesting** book.* ☐
c. *She's got such a **pretty** daughter.* ☒
d. *I saw a **young** girl waiting for a bus.* ☐
e. *I don't know.* ☐

1.

a. My brother **ran** a shop here when he was younger. ☐
b. Two men **robbed** a supermarket last night. ☐
c. Maria **managed** this project well last year. ☐
d. Hector **escaped** from prison five years ago. ☐
e. I don’t know. ☐

2.

a. I **ordered** a cheeseburger at Burger King. ☐
b. She **cooked** a very nice dinner for our guests. ☐
c. We **prepared** some snacks for the party. ☐
d. Who **ate** all the cheese sandwiches? ☐
e. I don’t know. ☐

3.

a. It was a **good** idea to come to the park today. ☐
b. My sister is having really **bad** luck this week. ☐
c. He was a very **powerful** man in the business world. ☐
d. There is a **strong** chance that I will get a bad mark. ☐
e. I don’t know. ☐

4.

a. She will **point** at the biggest diamond in the shop, I’m sure. ☐
b. Mom will **buy** some strawberries at the market today. ☐
c. Did your parents **give** you good advice for university? ☐
d. I don't really **pay** attention to the rings that women wear. ☐
e. I don’t know. ☐

5.

a. They **visited** Italy two weeks ago. ☐
b. I think I **went** deaf after the concert. ☐
c. He **became** a teacher after university. ☐

d. Alison **walked** to the metro alone. ☐

e. I don't know. ☐

6.

a. Jane **forgot** her keys at home. ☐

b. I **sold** my old laptop to a friend. ☐

c. She **left** her bag at the concert. ☐

d. I **threw** the flowers on the floor. ☐

e. I don't know. ☐

7.

a. I need to **cut** potatoes and onions. ☐

b. We **water** the flowers every day. ☐

c. I will **clean** the table before lunch. ☐

d. Can you please **wash** the carrots? ☐

e. I don't know. ☐

8.

a. This is a very **clever** thought. ☐

b. Mary was a very **intelligent** person. ☐

c. Leslie always wears **bright** colours. ☐

d. He gave me a **polite** answer. ☐

e. I don't know. ☐

9.

a. My brother likes **sweet** fruits. ☐

b. I met a very **nice** girl today. ☐

c. She is wearing a **short** skirt. ☐

d. Jessica bought a very **cute** dress. ☐

e. I don't know. ☐

10.

a. "Romeo and Juliet" is a very **sad** film. ☐

b. They had a very **unhappy** marriage. ☐

c. "Cats" is a very **boring** musical. ☐

d. A **strange** thing happened today. ☐

e. I don't know. ☐

11.

a. Can you **see** the sea from your hotel room? ☐

b. I can't **tell** the difference between these dresses. ☐

c. Pablo, you can't **say** this to your friend! ☐

d. I can **feel** the warm summer sun on my skin. ☐

e. I don't know. ☐

12.

a. Let's **begin** the meeting in five minutes.☐

b. Can you **introduce** me to your friends?☐

c. They can't **start** the party without me.☐

d. We will **organize** a conference next year.☐

e. I don't know.☐

13.

a. Who **drank** all the juice? ☐

b. Leo **chose** red wine for Saturday's dinner. ☐

c. I also **wanted** this chocolate cake! ☐

d. We've **selected** three candidates for this job. ☐

e. I don't know. ☐

14.

a. I think this trip to Italy is a totally **crazy** idea. ☐

b. Jennifer has had **bad** experiences with flatmates. ☐

c. Stop making **cheap** jokes about my friends, Gina! ☐

d. Our family always had a **poor** opinion of Uncle Tim. ☐

e. I don't know. ☐

15.

a. It was a very **hot**, dry summer in Texas. ☐
b. I hate the metro during the **rush** hour. ☐
c. He enjoyed drinking and driving **fast** cars. ☐
d. We met at a **busy** restaurant on Friday. ☐
e. I don't know. ☐

16.

a. Sam **got** lots of presents for his birthday. ☐
b. The new medicine **cured** him quickly. ☐
c. James **caugh**t a cold on holiday. ☐
d. My dad **fought** cancer for three years. ☐
e. I don't know. ☐

17.

a. Sally, you're in **deep** trouble now. ☐
b. Tom's new job is such a **big** achievement! ☐
c. I sat at the beach looking at the **dark** sea. ☐
d. You have very **low** standards for boyfriends. ☐
e. I don't know. ☐

18.

a. Have you already tried our **new** cream? ☐
b. You should eat more **healthy** food**.** ☐
c. It was a very **original** recipe. ☐
d. You need **fresh** tomatoes for a good gazpacho. ☐
e. I don't know. ☐

Task 4. *Look at the following sentences. In each case, choose the acceptable variant. In some cases both variants are acceptable, then you have to mark both of them. If you don't know either of the two expressions, mark "I don't know".*

1. You haven’t replied to my email.☐

   You haven't answered my email.☐

   I don’t know ☐

2. My small brother is a doctor.☐

   My little brother is a doctor.☐

   I don’t know ☐

3. I heard you have the flu, become well soon!☐

   I heard you have the flu, get well soon!☐

   I don’t know ☐

4. There were broken bits and pieces of glass on the floor.☐

   There were broken pieces and bits of glass on the floor.☐

   I don’t know ☐

5. I can't talk now, I'll call you back.☐

   I can't talk now, I’ll call you around.☐

   I don’t know ☐

6. My friend likes to mix and match different styles of furniture. ☐

   My friend likes to match and mix different styles of furniture.☐

   I don’t know ☐

7. The hotel also has a conference room.☐

   The hotel also has a meeting room.☐

   I don’t know ☐

8. I am sorry, I made a mistake. ☐

   I am sorry, I did a mistake. ☐

   I don't know ☐

9. Teacher and doctor are important professions. ☐

   Doctor and teacher are important professions. ☐

   I don't know ☐

10. The exam was so easy, piece of cake! ☐

    The exam was so easy, slice of cake! ☐

    I don't know ☐

11. Take it simple – don't get mad. ☐

    Take it easy – don't get mad. ☐

    I don't know ☐

12. They have a large house. ☐

    They have a huge house. ☐

    I don't know ☐

Task 5. *Look at the expressions in bold and imagine that you see them for the first time. Could you understand the whole expression if you knew the meanings of all the words in it? Choose "yes" or "no". If you are not sure about the meaning of the expressions, consult the translation provided in brackets below.*

1. There was a small **green apple** on the table.
   ☐ Yes
   ☐ No
   Translation: [manzana verde]

2. My daughter likes to **play with** her dolls in the garden.
   ☐ Yes
   ☐ No
   Translation: [jugar con]

3. I wanted to try bungee jumping, but **I got cold feet**.
   ☐ Yes
   ☐ No
   Translation: [me eché atrás]

4. - Can I have some water?
   - **Here you go!**
   ☐ Yes
   ☐ No
   Translation: [aqui tienes]

5. **Can you come closer,** I can't hear you!
   ☐ Yes
   ☐ No
   Translation: [puedes acercarte]

6. Look who is standing at the **church door**.
   ☐ Yes
   ☐ No
   Translation: [puerta de la iglesia]

7. Don't **give up**, you can do it!

   ☐ Yes
   ☐ No
   Translation: [rendirse]

8. **Are you busy?** I wanted to ask you something.
   ☐ Yes
   ☐ No
   Translation: [estas ocupado?]

9. I thought detectives were **heavy smokers**.
   ☐ Yes
   ☐ No
   Translation: [fumadores empedernidos].

10. Do you know what she **said to** him?
    ☐ Yes

☐ No
Translation: [dicho a]

11. - I get so annoyed with Steve!
    - **Tell me about it!** I think he is crazy.
    ☐ Yes
    ☐ No
    Translation: [Ya te digo!]

12. Dad was grilling chicken **in the back yard**.
    ☐ Yes
    ☐ No
    Translation: [en le patio de atrás]

13. **Come again**? Are you saying my mother gave you the money?
    ☐ Yes
    ☐ No
    Translation: [repitelo]

14. My friend Tommy died and I had to **break the news** to his family.
    ☐ Yes
    ☐ No
    Translation: [dar la noticia]

15. **I have already read the book**.
    ☐ Yes
    ☐ No
    Translation: [Ya he leído el libro]

16. He has **travelled around France** for a month.
    ☐ Yes
    ☐ No
    Translation: [viajado por Francia]

17. I went to **happy hour** at Belushi's bar last night**.**
    ☐ Yes
    ☐ No
    Translation: [hora con descuentos en un bar]

18. The party downstairs really **raised the roof** last night.
    ☐ Yes
    ☐ No
    Translation: [hizo mucho ruido]

19. I'm going to **clean your room** and bring you lunch.
    ☐ Yes
    ☐ No
    Translation: [limpiar tu habitación]

20. It is dangerous to **jump from** the bridge here.
    ☐ Yes
    ☐ No
    Translation: [saltar de]

21. Steven, can you **let the cat out** when you leave**?**
    ☐ Yes
    ☐ No
    Translation: [deja que el gato salga de la casa]

22. They're waiting for me to **kick the bucket** cause then they'll get more money.
    ☐ Yes
    ☐ No
    Translation: [morir]

23. We **ran out of** milk, can you go to the supermarket?
    ☐ Yes
    ☐ No
    Translation: [se ha acabado]

24. I am sorry, your flight **took off** half an hour ago.
    ☐ Yes
    ☐ No
    Translation: [despegó]

Appendix E. Collinearity diagnostics for continuous predictors

Variable	VIF
Pre-test	1.388
V_Size_1000	1.684
MetAwar	1.276
Stat1	1.142
Stat2	1.101
Age	1.445

Appendix F. The distribution of standardized residuals for GLM