The Said Construction:
Usage, Change, and Social Meaning in English and Spanish

by

Alicia Stevers

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2020

Doctoral Committee:

Professor Marlyse Baptista, Co-Chair
Associate Professor Ezra Keshet, Co-Chair
Assistant Professor Eric Acton, Eastern Michigan University
Professor Teresa Satterfield

Alicia Stevers

astevers@umich.edu

ORCID iD: 0000-0003-3751-5035

© Alicia D. Stevers 2020

DEDICATION

For my boys.

ACKNOWLEDGMENTS

As this project comes to a close, I find myself deeply grateful for the many different communities of support that have walked beside me through the past several years. I couldn't begin to thank everyone who has been part of this process. First, I owe the completion of this project to my committee chairs, Marlyse Baptista and Ezra Keshet. Marlyse gives me something to aspire to in every area of my life; she is a phenomenal scholar, a loving and involved mother, and one of the most kind and gracious people I have ever had the privilege to know. I could not have completed this project without her academic and emotional support. Ezra has also been there every step of the way; my first advisor and a person who has always believed my topic is worth the time and effort. He has gently challenged me and helped me figure out where I fit in this field.

I also express gratitude to Eric Acton, not only for being willing to take on this project and provide crucial feedback, but also for having paved the way for it. I couldn't be luckier that the linguist who's work best aligned with my interests happened to be right down the street and willing to be involved! I am also deeply grateful for Teresa Satterfield and her enthusiasm about my topic. She lent crucial support and insight for the Spanish portion of this project. I could not have done a cross-linguistic analysis without her native speaker insight and her affirmation that this type of analysis was interesting and worth pursuing in the first place.

The Linguistics department at Michigan is full of wonderful, brilliant people I am forever grateful to know. I would have quit long ago if not for the grace and kindness of Andries Coetzee. He has my loyalty forever! Jon Brennan reassured me early on that grad school is a great time to have babies, and to not let anyone tell me otherwise. He also opened his home to me and newborn Emmett for those early QRP planning meetings when I was sleep deprived and had zero cognitive function. Having a baby my first semester of grad school was not the easiest thing I have ever done, but our Michigan faculty and staff showed unwavering support as I tried (often seemingly unsuccessfully) to balance new motherhood with new grad studenthood. I also want to acknowledge Jen Nguyen, who was always there for a chat and would let me hang around her office way too long as I probably distracted her from way more important work. Will Styler provided key help with stats both in AA and SD, and I am so thankful for his expertise!

I also express gratitude to my SDSU thesis advisor, Jeff Kaplan, for being hands down the best teacher I've ever had and giving me confidence to pursue this project initially. It is not an exaggeration to say that he laid the foundation for this PhD and my career as a whole. I will feel like a success if I can spark even a hint of the enthusiasm for pragmatics he gave me, in my own students.

I am also fortunate to have been surrounded by so many amazing graduate student colleagues. Hayley Heaton, Rachel Weissler, and Yourdanis Sedarous, I don't have words for how much you mean to me! You three have been family to me and Dan, and the best aunties to our boys. Thank you for the support, shared holidays, memes, Marco Polos, and TV nights. Marjorie Herbert and Ian Calloway, my perfect little cohort,

you two are so special to me! Emily Sabo, Ariana Bancu, Rawan Bonais, Doms Bouavichith and Canning, Justin Craft, Tamarae Hildebrant, Marcus Berger, Jiesung Kim, Dave Ogden, Batia Snir, and all the other excellent grad students I've gotten to work and spend time with, thank you for the laughs and for making these years so bright.

Finally, my FBCA2 community, all the moments I thought maybe I'd stay in Ann Arbor forever were because of you. You are the type of church (and Paul and Stacey Simpson Duke, the type of pastors) I always hoped existed but seemed too good to be true until we found you. Cara Rosaen, thank you for being the friend I really needed In Ann Arbor. Our fall morning walks, music classes, and cozy tea-drinking couch sits were life giving, as was seeing our boys play together. My neighbors, Viki, Lisa, and Molly: you three made Ann Arbor feel like home. My cousins, Sean and Paula Lancaster, I'm SO glad I got to live closer to you and I loved every moment we got to spend together! Move to CA please. I am also forever grateful for Jazzercise (Ann Arbor and El Cajon!) for being my favorite place to dance and think, and for always giving me a confidence boost when I needed it!

I could not have done this without the support of my San Diego community. To have so many people who know and love you is truly special and I do not take it for granted. Regina, thank you for me laugh almost everyday from thousands of miles away. I love looking back on our 30+ years of friendship and celebrating how much we have accomplished; we're both living our childhood dreams and it's pretty darn great! Sara, the best long-distance friend a girl could ask for. Our near-daily MPs and shared memes keep me going. Vana, thank you for helping us move out to MI and for always

being just a text or Facetime away; I'm so thankful to have you as one of my people. Shannon, thank you for visiting me in MI twice! It means the world to me that you sacrificed time and money to fly across the country for some girl time!

Mom, even from across the country you still found ways to be the most involved, loving mom. Thank you for visits, care packages, sneaking Starbucks cards into my planner, and sitting on my couch 10 feet away to keep me focused. You are the best and I love you tuto mundo forever! Dad, thank you for always having confidence in me and telling me I am smart even when I didn't believe it. You always said you'd listen to my advice when I was a doctor. Guess what?

Emmett, If you remember the Michigan years at all, I hope you remember them fondly. I hope you remember being able to see the Michigan "M" from your bedroom window, fireflies, library trips, sledding, walking to the arcade, and running up and down the halls of the linguistics department eating snacks you snatched from my desk. Sunny, my little dissertation writing buddy, even though you are 100% a San Diego kid, this story is your story too and I'm so glad you were our ticket back to the sunshine. Dan, I don't even have words to express my love and gratitude for you. Thank you for moving across the country for me. These have been the hardest years, but also the sweetest. I couldn't have done it without you. You are the best partner and a constant source of support, strength, and reassurance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Abstract

This dissertation consists of three studies on the *said* construction (SC), characterized by a determiner-like use of ”*said*” followed by a noun (N2), typically given (in some sense) and licensed by an antecedent noun (N1).

> (1). When arguing a point with **an opponent**$_{N1}$, she was accustomed to swift deference (provided that **said opponent**$_{N2}$ was male, straight, and his eyes worked). [Freisner 2011]

The first study presents a corpus-based analysis of SC, showing how the construction has changed over time in form and genre of usage. Early tokens of SC are likely to cooccur with an additional definite determiner, as in “*the said* N” whereas more recent tokens are more likely to lack this additional determiner. I also show evidence that this once formal construction (commonly used in legal discourse) has shifted into informal genres of communication such as television and web-based discourse. This modern usage also seems to be accompanied by a change in meaning, which leads to the second study; this study presents the results of a social media-based experiment designed to answer questions related to the social meaning of SC. Participants of this study were significantly more likely to react to SC-containing sentences with a “haha” reaction, as opposed to a standard “like” reaction. Finally, I present a cross linguistic analysis of a similar construction in Spanish, the *dicho* construction. This study reveals distinct differences between the usage and meaning of “*said”* and “*dicho*”; *dicho* is significantly more likely than *said* to be used with inferred information and significantly less likely to be used with an explicitly mentioned antecedent. Furthermore, a parallel

Spanish social meaning experiment shows that Spanish speaking participants are not significantly more likely to react to a *dicho*-containing sentence with a “haha” reaction than a standard one. Results of these Spanish studies show that seemingly similar constructions can follow distinct developmental paths, leading to modern usages that are different in usage and interpretation.

CHAPTER 1

Introduction and Theoretical Background

1.1 Introduction

This dissertation investigates the *said* construction (SC), a standard English construction, usually characterized by the use of *said* in place of a determiner, followed by a noun (N2), typically given (in some sense) and licensed by an antecedent noun (N1).

> (1). When arguing a point with **an opponent**$_{N1}$, she was accustomed to swift deference (provided that **said opponent**$_{N2}$ was male, straight, and his eyes worked). [Freisner 2011]

This work will present the results from three studies on SC, using corpus based and experimental methods. These studies are centered around the following research questions:

1. What is the behavior and grammatical status of the *said* construction, and how has it changed over time?
2. What is a speaker's motivation for choosing to use this construction over an alternative, more standard determiner form, and how do addressees interpret its meaning or speaker's intentions when they encounter it?

3. How does this construction compare to the similar *dicho* construction present in Spanish, and do they have the same social meaning in the two languages?

While there is no conventionally agreed upon definition of givenness, studying SC informs theories of givenness by having a distinct role within a discourse. Chapters 1 and 2 discuss some existing theories of givenness and propose that SC fits best with a linguistically based model rather than one that includes inferred information or world knowledge. Closely related to givenness, SC has implications in existing theories of information status, determiners and definiteness. Through this project, I also show that a historical change in the form of a construction can be related to a change in social meaning. In English, SC has undergone a change in form and has adopted a unique social meaning. In Spanish, on the other hand, this structural change has not taken place, nor is there a social meaning attached to this construction.

The first study uses corpora to show the construction's change in form and discourse function over time. I show that early tokens of SC are more likely to use an additional determiner, taking the form *the said N*, while also showing evidence that over time this additional determiner has been dropped, with current tokens favoring a usage that takes the form of *said N*. This study is also evidence of a shift in the types of genres SC appears in, from formal forms of discourse such as legal and academic, to informal such as television, spoken language, and (crucially) web-based discourse. Finally, using these aforementioned changes as evidence, I argue that *said* when used in this construction displays determiner-like properties, and as a determiner, has a relationship to the information that precedes it that is unique and distinct from the relationships of

other determiners.

Using the findings of the first study as a starting point, the second study seeks to understand why a speaker would choose to use SC as a determiner when there are other, more standard determiner forms available. I hypothesize that this has to do with a potential social meaning carried by *said* that is unique to its usage and distinct from other determiners. Therefore, Study 2 uses experimental methods to show that SC can be used to convey meanings of humor and mock-intelligence, and that these meanings are distinct from any potential meanings conveyed by other English determiners. 200 native English speaking participants were presented with sentences containing either SC or other standard determiner forms such as *the* and *that*. Participants were asked to read and react to these sentences using an emoji-based reactions schema, designed to resemble the reactions bar used on social media sites such as Facebook. A general linear model was used to calculate correlations between participant reactions and determiner types, showing that sentences containing SC were significantly more likely to be regarded as humorous from participants than identical sentences that contained a standard determiner [$p<.001$]. Additionally, participants were asked a series of open-ended questions about perceived usages and meanings of SC. Results of this task showed that study participants believe individuals use this construction when they wish to be perceived as humorous or intelligent. The results of this experiment provide evidence that SC has a unique social meaning, and that English speakers may choose to use this construction when they wish to convey a broad range of social meanings in discourse such as humor, intelligence, and ironic formality.

Finally, the third study I present attempts to construct a parallel analysis of SC's sister construction present in Spanish, which I call the *dicho* construction (DC), as shown below:

> (2). Vamos a hablar de Cáceres: **dicha** ciudad fue construida en 25BC.
> *And now we come to Caceres: said city was built in 25 BC.*

This study shows through corpus based and experimental methods that *dicho* is used differently from SC in regard to the types of information with which it is most likely to occur; *dicho* is used much more frequently with inferred information than SC, which favors an antecedent that has been explicitly mentioned. Furthermore, a social meaning experiment administered to native Spanish speakers shows that they do not consider *dicho* to be humorous, and shows no significant effect between determiner type; identical sentences containing either *dicho* or a standard determiner such as *ese* or *eso* were equally as likely to be considered humorous or non-humorous. I believe this can partially be explained by the idea of un/expectedness, which is discussed below in section 1.5.1.

The results of these studies all relate back to the notion that this use of *said* in English has undergone a significant change in form and function, and it is now most commonly used informally by speakers who wish to convey a subtle meaning of humor and/or mock-formality. In Spanish, however, this construction is still used more formally, and does not have an apparent social meaning.

More broadly, this project contributes to research on definiteness, determiners, and social meaning (Hawkins 1978, 1991, Prince 1988, 1992, Birner & Ward 1994,

Ward & Birner 1995, Abbott 1997, Campbell-Kibler 1999, Acton 2014, 2019, Acton and Potts 2014, inter alia) supporting past research that determiners can convey distinct meaning in a discourse in spite of the fact that they are a small, closed class of words. Furthermore, *said*'s semantic shift from formal to informal and a potential syntactic shift from adjective position to determiner position shows that determiners are not immune to linguistic change. The Spanish work shows that this construction has not (yet) undergone the same change in Spanish that it has in English, and that it is used and interpreted differently by speakers of the two languages.

The rest of this chapter will provide an overview of relevant literature and theory for this project, for the purposes of situating *said* within existing literature and providing a framework for the rest of the dissertation. While the bulk of the analysis will be in chapters 2-4, this chapter will lay a foundation for analysis by tying in and relating *said* to the relevant literature.

1.2 Broad introduction to the relevant literature

Over the past several years, renewed attention has been given to the intersection of sociolinguistics with pragmatics, building on the notion that language can have distinct meanings and social implications within different discourse communities, and that this meaning can be distinct from that which can be interpreted strictly from the semantics of an utterance itself. It is well known that language contributes to the construction and performance of social identities across discourse communities, but sociopragmatics aims to investigate pragmatic meaning of language in specific and

more local contexts of language use (Leech 1983:10) The analyses presented in this dissertation extend over several subfields and have implications for several long standing theories: The corpus-based analysis discussed in chapter 2 examines the diachronic change and grammatical status of *said*, which relates to existing work on information status, specifically definiteness, determiners, and givenness. Chapter 3 then presents results of a social meaning experiment that draws from existing literature in sociolinguistics, social meaning, and how these ideas relate to determiners and demonstratives. Since the *said* construction has not been formally studied before, one of the objectives of this dissertation is to lay a foundation for future research on this and similar constructions, both in English and in other languages like Spanish.

To establish a framework in which to anchor the studies presented here, I draw from scholarly works on givenness, definiteness, determiners/demonstratives, and social meaning/sociolinguistics. The literature on determiners and definiteness is particularly important because I argue that *said*, when used in this construction, behaves like a determiner in terms of how it interacts with other information that precedes it. Furthermore, it seems that *said* patterns most similarly to definite determiners, in that it is most commonly used with information that has already been mentioned or established in the discourse. The literature I present here on determiners and definiteness provides detailed analyses and descriptions of other definite determiner constructions, which gives this dissertation not only a framework for analysis, but also relevant information on these various determiner constructions, providing a starting point for showing how *said* is similar to and different from

determiners like *the*. The following literature review presents and summarizes important scholarship on these topics that will be foundational in constructing a holistic framework for this project.

1.3 Determiners and in/definiteness

This section discusses the notions of definiteness and indefiniteness and how they relate to the identifiability of information[1] within a discourse. Before beginning this section, I want to acknowledge that different languages express definiteness differently; it can be expressed via a separate word, such as *the* in English or *el* in Spanish, an affix, as the suffix -et/-en in Norwegian, or can even be denoted by word order, as in Russian. For the majority of this project, I am focusing only on English and use the term *definiteness* to refer to notions of information status and the types of determiners that are used in various informational contexts. It is generally observed (with many noted exceptions, several of which will be discussed in this chapter) that entities which a speaker believes a hearer can identify are *definite*, and these types of nouns co-occur with *the*. Lyons (1999:3) provides many examples of these types of usages, distinguishing between what makes an entity definite in various contexts. Situationally, one could ask a family member (who lives with the speaker in a house with more than one bathroom) to “Put these clean towels in the bathroom, please”. (Lyons 1999 example 3). In this case, the speaker believes the hearer can identify the bathroom due

[1] I will define *information* as any entity linguistically evoked in a discourse or salient via extralinguistic context. Much of this work is rooted in literature on information status, so I use the term *information* in congruence with that literature.

to his or her familiarity with the house, and the knowledge that these clean towels belong in a specific bathroom (due to color, for example) as opposed to any other bathroom(s) in the house. This context and identifiability make the speaker's use of *the* a likely choice. Entities that are not identifiable or situationally unique are preceded with indefinite *a*. A speaker might say, "I tried *a* new restaurant yesterday", using *a* since he or she does not believe the hearer can identify the restaurant being mentioned. Words like *the* and *a* are considered determiners, which function to modify the nouns they precede. The word "determiner" can be used to refer both to the type of word, as well as its position syntactically (Lyons 1999). For the sake of this dissertation, I use the term to refer to the words in that word class, such as *the*, *a*, *that*, *this*, etc. Any reference made to syntax will use the term "determiner position", specifying the location in the syntactic structure.

Despite decades of research on the distributions of these determiners and the types of information with which they most likely occur (Haviland & Clark 1974, Li & Thompson 1976, Prince 1988, 1992, Comrie 1989, Ward & Birner 1995, inter alia) there is still not a single theory or description of definiteness that can fully account for the distribution of the definite article in English. One of the objectives of this dissertation is to introduce *said* into the conversation, showing that it behaves similarly to a definite determiner but with a different distribution in terms of the types of information with which it occurs. The following sections provide an expanded discussion of some existing theories of definiteness and determiners, so that *said* may be situated within the literature.

1.3.1 Referential expressions: Gundel, Hedberg & Zacharski

While individual determiners are usually good indicators of definiteness, it is important to look at the entire determiner phrase in which the determiner is used. These types of determiner phrases that are used to indicate a discourse entity are often called referring/referential expressions (Lyons 1977). In Stevers 2014 & 2017 I argue that *said*, when used in SC as part of a referential expression, displays determiner-like properties due to its syntactic position and the way it interacts with preceding information in the discourse. If it can be considered a determiner, it is certainly non-canonical and must have constraints of usage, such as where it can occur syntactically, the types of information with which it is most likely to be used (whether old, new, or inferred) and the types of antecedents that precede it. One of the main questions this project seeks to answer is why speakers would choose to use SC when there are other perfectly suitable and arguably more standard forms readily available to them, such as *the* or *that*. Gundel, Hedberg, and Zacharski's (henceforth "GHZ") 1993 paper pursues a similar question, asking what knowledge enables a speaker to select a particular form from other viable alternatives, and similarly, what knowledge enables a hearer to correctly identify a referent if one is not explicitly clear from the referential expression used by the speaker (GHZ 1993). The authors introduce the idea of cognitive status, which they define as information about location in memory and attention state; in other words, cognitive status describes what is in a hearer's mind at the time of utterance. This could be an exact mental representation of the item under discussion, or a type/generic representation (perhaps the hearer's own cat versus the idea or concept of some

unknown cat somewhere else). It could also be something known as a result of general cultural knowledge. For example, a speaker could assume that a hearer is familiar with concepts relating to the Civil War even if they had never discussed it together before (and certainly neither were present at the actual events). This assumption would place *the Civil War* on a different cognitive level than that of a completely unheard of entity that was not in the hearer's awareness or knowledge at the time of utterance. These cognitive statuses can also come from direct linguistic input from the current conversation or even past dialogues between speaker and addressee, re-referencing discourse entities that have already been discussed by participants.

GHZ suggest that individual determiners activate different cognitive statuses in the mind of the hearer, each of which allow for a wider or narrower set of possible referents. This means that in any given conversation, the speaker has an idea of the ability of their addressee to understand a particular utterance, and this knowledge guides the speaker's choice of determiner. For example, if the speaker is referring to something they just experienced and that they believe the hearer is unfamiliar with, they will choose an indefinite determiner "a". For example:

(3). Speaker A, to B: "I tried *a* new recipe yesterday."

Speaker A does not assume B is familiar with this recipe, and perhaps this is the first mention of any kind of cooking the two have discussed up to the time of utterance. Alternatively, if the speaker intends to discuss something he or she has discussed or experienced with this particular addressee on another occasion, they will choose "the" or "that N". Returning to the above example, imagine the two speakers had previously

discussed the possibility of making a cheesecake from A's grandmother's recipe. Speaker A might say something such as:

(4). I finally tried the/that recipe! It was delicious!"

Speaker A assumes B recalls the past conversation; from the perspective of the hearer, the determiner that they hear provides clues about what their level of awareness about the referent should be. An "a" provides a clue that the referent is new at least to the present conversation, whereas a definite determiner such as "the" informs the hearer that they should be aware of the referent being mentioned.

To construct their analysis, GHZ introduce a six-level givenness hierarchy based on the idea of cognitive status, shown below in Figure 1. This hierarchy is a useful tool in beginning to understand determiners and how they refer to preceding information in a discourse, as it demonstrates the level of awareness a speaker should have about an entity in the discourse in order to felicitously use each common referential expression. In terms of SC, the hierarchy provides a starting point from which to consider how the construction is used in terms of the types of information with which it can co-occur. GHZ explain the appropriate distribution of their six expressions, showing how each can only be utilized only when the necessary and sufficient cognitive status for use is met.

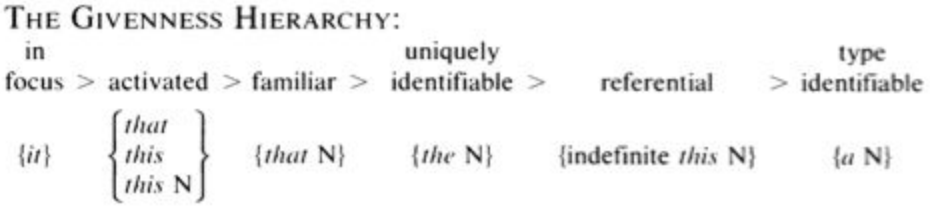

Figure 1.1: (GHZ hierarchy): Givenness Hierarchy, Gundel, Hedberg & Zacharski 1993

The authors define each cognitive status and provide the following examples (moving right to left):

Type Identifiable: The addressee is able to access a representation of the type of object described by the expression. (Marked by the use of indefinite *a*).

(5) *I couldn't sleep last night. A dog (next door) kept me awake.*

Referential: The speaker intends to refer to a particular object or objects. (Marked by use of *this N*)

(6) *I couldn't sleep last night. This dog (next door) kept me awake.*

Uniquely identifiable: The addressee can identify the speaker's intended referent on the basis of the nominal alone. (Marked by use of definite *the*).

(7) *I couldn't sleep last night. The dog (next door) kept me awake.*

Familiar: The addressee is able to uniquely identify the intended referent because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has.)

(8) *I couldn't sleep last night. That dog (next door) kept me awake.*

Activated: The referent is stored in the hearer's immediate, short-term memory, either through having been retrieved from long-term memory or from being salient in the current discourse, either linguistically or extralinguistically (immediate situational context). This status accounts for use of pronouns and definite demonstrative *this*.

(9) *I couldn't sleep last night. That/this kept me awake.*

(10) *My neighbor has a dog. That dog kept me awake last night.*

In Focus: The referent is currently salient in the discourse, as the center of attention. It most likely involves important information from most recent utterances in the discourse.

> (11) *My neighbor's dog bit a girl. It's the same dog that bit Mary last summer.*

This paper sets a foundation for an analysis of *said* as a determiner because it effectively distinguishes between various determiner forms and the types of information with which they are most felicitously used. Within the boundaries of a conversation between speakers, it is logical to assume that the speaker of any/each of the above sentences would pick a referential expression based on how much he or she believes the hearer knows about the intended referent of their sentence. One would not choose to say, "It kept me awake last night" if the dog had not already been established as a salient referent in the dialogue. He would instead use "A dog kept me awake last night," or the false (non-deictic) definite, "This dog kept me awake last night." As will be shown in chapter 2, *said*, like several of the other determiners in the GHZ hierarchy, is most likely used with a noun that has already been brought into the conversational space. In other words, it cannot behave like an indefinite determiner, introducing new information into the discourse.

These various referential expressions and levels of awareness make up GHZ's idea of *cognitive status*. Understanding these levels of awareness is helpful in beginning to theorize where determiner *said* fits among other determiner types and what level of awareness the speaker may believe the hearer should have if he or she were to choose to use it in a discourse. Furthermore, *said* seems to behave most like referential

expressions towards the left side of the GHZ hierarchy (such as those under the categories of *in focus*, *activated*, and *familiar*), seemingly satisfying higher cognitive statuses like the ones achieved by the use of definite determiners. The GHZ hierarchy provides insights into how the distribution of referential expressions is closely related to the knowledge which the speaker assumes the hearer possesses, and provides a starting point from which to begin thinking about *said*.

1.3.2 Definiteness and information status

Other research has constructed theories of definiteness based on similar ideas: that definite referents must be distinguishable and identifiable from other possible referents, even if they are not specifically known (Hawkins 1978, 1991, Holmback 1984, Abbott 1993). There have been arguments for theories of definiteness and determiners that are more rooted in the status of information within the discourse than the cognitive status of the hearer or his or her ability to identify a particular referent. Ward and Birner (1995), for example, challenge the familiar notion that definite determiners are only used with information that has already been stated or established in the discourse, while indefinite determiners are used to introduce new information. This analysis is rooted in work by Prince (1992), who was the first to offer a distinction between hearer status and discourse status. Prince asserts that information can be old or new to either the hearer or the discourse; hearer status has to do with what the speaker believes to be true about the hearer's knowledge, while discourse status relates to whether or not an entity has been mentioned in the present discourse. For example, if speaker A saw a fox in his yard, he would not say to his addressee with no prior information, "I saw the/that fox

in my yard today," but would instead use the indefinite determiner. However, if the discourse participants had seen this specific fox together on another occasion, using *the* or *that* would be acceptable. In this case, *the/that fox* is hearer old according to Prince 1992.

Operating under Prince's framework, Ward and Birner argue that "the use of a definite NP is said to require that its referent be known, given, or inferrable in context" (Ward and Birner 1995, pg 724) and provide evidence that a definite determiner can be used with a referent that is not necessarily uniquely identifiable to the hearer, or even one that has been previously mentioned in the discourse at all. For instance, they show that a speaker can say, "It's hot in here. Could you please open the window?" even if there are multiple, equally salient windows. Additionally a speaker could encourage an interlocutor to take "the elevator" to a higher floor, even if there are multiple elevators available and ready for use (Birner and Ward 1994, their examples (2a) and (2b), pg 1). These examples highlight the problem that there is still no single theory of definiteness that can fully account for the distribution of the definite article in English. This also shows that there are underlying factors that contribute to a speaker's choice of referential expression, and that there are not clean-cut lines between the types of information that can be referenced by any single determiner. Certain determiners are used more felicitously in certain contexts, but these contexts are dependent on a variety of factors such as the knowledge states (both knowledge of the conversation and common world knowledge) of each discourse participant, and the extralinguistic context provided by the conversational setting. Based on my corpus-based collection of tokens

(which will be discussed more in depth in chapter 2), it seems that SC can only be used to refer to an entity that has already been mentioned in the discourse (Hearer-old, Discourse-old, under Prince's model), and there do not seem to be the same types of exceptions that we see with *the*, as discussed by Birner & Ward (1994). Tying this back to GHZ's analysis and taking *said*'s distribution into account, it seems that it activates a cognitive status very similar to GHZ's *familiar* category, but with the crucial difference that the referent used with *said* must be in the recent short term memory of the hearer due to a recent mention, and not in long term memory, as allowed by GHZ's definition of this category.

1.3.3 Definiteness and existential-there sentences

One type of sentence that is especially problematic to existing theories of definiteness are *existential-there* sentences (also called *there*-sentences), in which the ability of *there*+BE to precede definite NPs has been contended with in the literature; many authors have stated that this cannot be done (Guéron 1980, Jenkins 1975, Milsark 1977), while others have provided evidence of sentence constructions where it is possible (Abbott 1993, Lakoff 1987, Prince 1988 & 1992, Birner & Ward 1995). Many studies on *there*-sentences confront issues of speaker and hearer knowledge, and of what should be considered old or given in a discourse; these are concepts that are directly related to this work on SC, and arguments made in papers on *there*-sentences provide relevant points for constructing my SC analysis. For example, studies on existential *there* (to be called *ex-there*) have noted the tendency of *there*-sentences to contain forms of the verb "be" followed by indefinite NPs, while equally as many authors

have noted clear exceptions (Rando & Napoli 1978, Guéron 1980, Freeze 1992, inter alia). Consider the following examples:

> (12a) We walked into an Italian restaurant and were seated immediately. **There was a dead fly** on the corner of our table.
>
> (12b) *We walked into an Italian restaurant and were seated immediately. **There was the dead fly** on the corner of our table.

As demonstrated in the examples (12a) and (12b), the use of an existential-there clause lends most naturally to an indefinite referent, and logically so; bringing something into existence in the conversational space (the natural use of ex-there) would by default mean that it is hearer and discourse new, and should be introduced with an indefinite determiner. Additionally, these types of *ex-there* sentences are not permitted with *said* as determiner in the NP:

> (12c) *We walked into an Italian restaurant and were seated immediately. **There was said dead fly** on the corner of our table.

Once the fly has been mentioned or established in the discourse, any recurring references to it can then be used with a definite determiner or with SC, but at first mention, use of *the* or *said* are infelicitous.

According to some later research, however, it is possible to come up with counter-examples; there are certain types of sentences in which ex-there can be used with a definite referent. Ward and Birner 1995 provide an analysis of these types of sentences, and assert that while the NP following a verb in a *there*-sentence must be (or, crucially, “behave as”) Hearer-new in accordance with Prince’s (1992) analysis, definiteness merely requires that the speaker believes the addressee can identify the

referent. In other words, it is *uniquely identifiable* in accordance with GHZ's analysis. Consider example (13) (Ward and Birner 1995, example 12, taken from Prince 1992, example 5):

(13a) There were the same people at both conferences.
(13b) There was the usual crowd at the beach.

In these examples, one can assume that the speaker believed the hearer could create a valid mental representation of these referents, so although the construction uses a definite determiner in conjunction with *ex-there*, Prince argues that the NP is still behaving as Hearer-new. This is not a possible behavior of *said* in SC. In my corpus analyses and through collecting *said* data "in the wild"[2], I have not found any instances of *said* behaving in a way that is Hearer-new-like as does *the* in example (13); rather any sentences that use both *there* and *said* would have to be deictic, pointing to (physically and/or linguistically) a referent that had already been established in the discourse. The lack of these types of *ex-there* sentences with SC further points to the distribution gap between *said* and other determiners, highlighting the need for a complete analysis of SC and its relationship to information in a discourse.

Abbott (1997) more closely examines these types of sentences and the problematic cases they pose for theories of definiteness and givenness. According to Abbott, Ward's and Birner's analysis fails to address some major issues, and she asserts that all previous theories of information status are incomplete in that they do not account for all types of *there*-sentence data. Abbott makes a crucial point that an

[2] I use this term to refer to tokens I have come across in my own media usage (social media, television, email, books, etc.), or tokens that have been sent to me over the years by colleagues.

addressee's immediate consciousness and their permanent memory store should be handled differently in information status research, and uses the GHZ hierarchy as evidence to support this assertion. This ties into the present research; SC seems to be more dependent on the hearer's immediate consciousness, whereas standard demonstrative determiners can probe a hearer's memory store. It is not felicitous to use SC to reintroduce a referent that was established in a previous discourse; examples of this sort are not used or attested in any of the hundreds of SC tokens collected. Consider the following examples:

(14a) A: Did you end up trying that new Thai food place we saw?
B: Yes and said Thai place was worth the wait.
(14b) *B (to A, with no established referent in the present discourse):
Guess what! Said Thai place we saw was delicious!
(14c). B (to A, with no established referent in the present discourse):
Guess what! That Thai place we saw was delicious!

In (14a), speaker A establishes a new restaurant as the referent, which B then refers to using SC. In (14b), however, B infelicitously mentions the Thai place using SC, although it is a referent that both speakers are familiar with. In this context, a standard demonstrative, such as in (14c), would have been felicitous. This distinction is helpful when analyzing tokens of SC, and can partially help explain the relative inability of *said* to be used to refer to inferred information. Understanding how certain determiners are allowed or disallowed in this type of sentence further clarifies *said*'s relationship with information it appears with. As shown above, the use of *said* in this type of sentence

seems awkward, furthering the divide between *said* and other determiners distributionally.

1.3.4 Demonstratives and generics

Also relevant to the present project is literature on demonstratives and generics, which provides evidence that, while similar to demonstratives in the types of antecedents it can follow, *said* has a different relationship to these antecedents than standard demonstratives do. Authors have shown that demonstratives, like other definite determiners, can be used generically to refer to kinds of entities instead of specific instances (Bowdle & Ward 1995). Consider example (15) (Bowdle and Ward, example 13):

(15) A: My dog was attacked by a porcupine yesterday.

B: Those porcupines are very territorial.

In example (15), the demonstrative *those* is not referring to any specific set of porcupines, but rather the entire species as a kind. The authors argue, however, that this type of generic demonstrative usage is not felicitous with antecedents that are more taxonomically basic, under the tri-level categorization of taxonomies, presented in Rosch et al. 1976. Under Rosch's classification, entities are grouped according to levels of similarity. In his most general category, *superordinate*, one would find very broad categories such as "animals" or "vehicles". Members within these types of categories could potentially be very different from one another. For example, within the category of "animals", one would find members such as "birds", "reptiles", and "dogs". The next most specific category is "basic", which efficiently groups together members of more

specific types, such as (all) “birds”. Rosch’s most specific level of categorization is “subordinate”, which includes a single type of each category. In the case of birds, this would be a single species, such as “cockatoo.” While the basic category of birds would include quite a bit of variation (such as across different species), one can expect to find significantly less variation within the subordinate category, as (to my non-expert knowledge) cockatoos generally look like and behave similarly to other cockatoos. Returning to Bowdle and Ward’s analysis, just the same, the category of ‘porcupine’ is not divisible into further subordinate categories. This, however, is not the case with other basic categories such as ‘dog’ or ‘cat’, which can be further divided into specific breeds, e.g. German shepherd or Siamese. This distinction in taxonomic category of information, Bowdle and Ward argue, can account for a difference in the felicitous usage of generic demonstratives. Consider the following examples, in which the use of a generic demonstrative is felicitous only for (16b) and (17b), which have referents that are more specific than those of (16a) and (17a) (Bowdle & Ward, examples 9 and 10):

(16a)A. My roommate just bought a dog.
Ba. Dogs make great pets.
Bb. #Those dogs make great pets.

(16b)A. My roommate just bought a labrador.
Ba. Labradors make great pets.
Bb. Those labradors make great pets.

(17a)A. I’m thinking about buying a car.
Ba. Cars can be expensive.
Bb. #Those cars can be expensive.

(17b)A. I’m thinking about buying a sportscar.
Ba. Sportscars can be expensive.
Bb. Those sportscars can be expensive.

In examples (16a) and (17a), “dog” and “car” are both taxonomically superordinate. There is no assumption that members included in the set of all dogs or the set of all cars are homogenous. Bowdle and Ward argue that at this superordinate level, there exist too many internally contrastive sets within each category, making vast generalizability difficult and infelicitous. In (16b) and (17b), however, neither “labradors” nor “sportscars” are taxonomically superordinate; both would be considered subordinate, or specific enough that they are not drastically subdividable into further sets. This allows speakers to use a generic demonstrative, as it is not surprising that in general, labradors make great pets and sportscars can be expensive.

Said, although closely behaving like demonstratives in usage and distribution, does not seem to have this generic usage. The use of *said* in any of the above examples would imply reference to one single, identifiable entity:

(17c)A. I’m thinking about buying a sportscar.
Ba. Sportscars can be expensive.
Bb. Those sportscars can be expensive.
Bc. #Said sportscars can be expensive.

This distinction shows that *said* has a different relationship to the information that precedes it than other demonstrative determiners. While standard demonstratives can be used with generic NPs, this does not seem to be possible with SC. In other words, generic demonstratives can represent “kinds”, whereas *said* seems to have to refer to the exact entity mentioned, not just a general representation. There do, however, seem to be cases in which *said* can potentially be used to refer to kinds, but the specific kind

must be established in the discourse before use of *said* and is most felicitous if a subtype/type relationship is formed:

> ?(18) The cockatoo, native to Australia, is known for its crested feathers. Said bird can also make a great pet.

In example (18) above, the generic use of *cockatoo* establishes *bird*, making the use of a *said* DP felicitous, as its referent is at that point Hearer-old according to Prince 1992.

1.4 Pragmatics: Grice (1975)

Also foundational to this project are classic works in pragmatics, as a significant part of this dissertation is dedicated to understanding what is communicated and understood by discourse participants through the use of SC. Grice (1975) introduced the *Cooperative Principle* and the conversational maxims that accompany it. The idea behind the Cooperative Principle is that there are general conversational guidelines that speakers are expected to observe, and failure to do so marks the speaker as uncooperative. Specifically, this principle states that one should "make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice 1975, page 45). This principle is composed of the maxims of Quantity, Quality, Relation, and Manner and the idea is that when these maxims are observed, the result should be a fully effective and maximally informative conversational exchange (Grice 1975).

Specifically relevant to the present research are the maxims of Quantity, Relation, and Manner. The maxim of Quantity states that one should make his or her

conversational contribution as informative as is required. For example, a surgeon who asks her assistant for a scalpel expects to be handed only what she asks for; receiving a scalpel, forceps, and gauze would not be maximally helpful in that moment. In conversation, stating more or less information than is required for one's specific conversational turn is a violation of the maxim of Quantity. The maxim of Relation states that one should not say something which is inappropriate or unrelated to the conversation; the surgeon would not expect to receive a puppy if she asked for a scalpel. In conversation, stating something not related to the present conversation would violate this maxim of Relation. The maxim of Manner has several sub-maxims; a cooperative speaker is to avoid obscurity and ambiguity, be as brief as is necessary for the speech act, and be orderly in how information is presented (Grice 1975).

In summary, when operating under Grice's Cooperative Principle, the violation of a conversational maxim will lead a hearer to draw an inference about the meaning of the utterance for the purpose of maintaining the assumption that a speaker is being maximally informative and cooperative. In terms of SC, if we assume that a speaker is observing the maxims and making an effort to abide by the Cooperative Principle, the speaker's use of a nonstandard form like *said* suggests that the speaker believes that *said* contributes appropriate, relevant information to the discourse that an alternative determiner would not. The choice to use a marked form must be worth the conversational cost to the conversation. In the case of *said*, it may seem uncooperative because its use can be argued to violate manner, as it is a less likely choice when considered against other determiners. A hearer, therefore, would have to infer that *said*

is serving a purpose in the discourse in order to hold the assumption that their fellow interlocutor is being cooperative. Findings presented in this project show that this assumed purpose could be related to social meaning.

1.5 Social meaning

When considering *said* in light of Grice's maxims and Cooperative Principle, it may be the case that SC would be maximally informative in some conversations and less so in others. A growing body of work is in the area of semantic and pragmatic meaning within social contexts. Acton (2014) develops a framework for analyzing these types of phenomena, which will be used in my analysis of the social meaning of SC presented in chapter 3. The purpose of this framework is to bring together meaning-based research in several subfields (namely semantics, pragmatics, and sociolinguistics) in order to address questions that have traditionally not been asked: ones that examine meaning and variation within and across the context of specific social groups.

1.5.1 Acton: A framework for analyzing social meaning and determiners

Acton's framework has four main principles which are built upon the Gricean maxims and operate under the assumptions that speakers have conversational goals, meanings underlying their contributions, and expectations for how a given conversation will unfold. The first principle underlying Acton's framework is the Associations and Entailments Principle (AE), which states that the context of a speech act includes all items entailed by it and associated with it. This principle addresses the truth that within

the context of a language or even a smaller community of practice within a larger language group, words or phrases have meanings associated with them. The words “thin,” “gangly,” and “underweight” are all terms that could be referring to the same general size, but each term carries underlying connotations or associations, whether positive or negative. I argue that this principle is equally applicable to determiners, as a speaker’s choice of determiner provides clues to the hearer about the type of knowledge they should be accessing to understand a particular referent. For example, *the* and *that* can both refer to a definite referent, but a speaker will select the one they believe is most appropriate based on what they believe the addressee knows at the time of utterance. Furthermore, *that* and *said* can often be used in similar linguistic contexts, but I will show that a speaker who uses *said* will have specific conversational goals that can best be achieved through the use of this particular construction.

The second principle of the framework is the Full Significance Principle (FS), which states that the significance of an utterance is determined by context, and also how that utterance is different from other related utterances that could potentially be selected as alternatives. Under this principle, the relevant set of utterances available to discourse participants at any given moment is determined by speaker and hearer expectations about the discourse, their beliefs about one another, and their understanding of the immediate extralinguistic context of the conversation. This speaks to the assumption that the scope of shared knowledge between speaker and hearer is quite large, covering everything from immediate situational knowledge to personal history between discourse participants. The purpose of this principle is to motivate

comparisons between related linguistic units. This means, In terms of SC, that the FS principle is foundational for warranting the comparison between *said* and other relevant determiners that are equally accessible to the speaker at the time of utterance. It ties together conversation context with the different conversational goals that could be accomplished by any possible determiner in light of that context. For example, depending on the conversation leading up to a particular utterance, as well as the conversational setting/extralinguistic context and believed shared knowledge with the hearer, a speaker will decide if a definite or indefinite determiner is more appropriate to accomplish his or her next conversational goal. In the case of *said*, a speaker's decision to select this form is less dependent upon situational context and more so the context of what has been mentioned in the discourse and perhaps even underlying intentions of conveying specific types of social meaning.

The third principle, called the Differential Importance of Different Alternatives Principle (DI), states that alternative utterances have different levels of importance, which is dependent both on how related they are to the original utterance and how much they align with the conversational expectations of the discourse participants. The framework goes on to say that an expected response is more likely to play a role in the interpretation of its antecedent; it is more likely that inferences will be drawn from expected and related utterances than seemingly irrelevant ones. The main idea of this principle is that in any given context, some alternatives will be more acceptable than others, and that not all alternatives have the same level of importance to the discourse or serve the same purpose. This is less applicable to the case of SC because with

determiner forms, it does not seem as likely that conversational context plays as significant a role in determining a set of possible alternatives; rather it makes sense that at any given time (within the constraints of information status), speakers will have the same set of alternatives available to them with little variation. This is due to the fact that, even though in any given context some determiners may be more appropriate than others, determiners are a small, closed word class with the same relatively limited set of alternatives available at any given time (*the*, *a*, *that*, *this*, *these*, *those*, pronouns, numbers and quantifiers such as *some*, *every*, *many*). Many more alternatives are available in large, open classes like nouns, with hundreds of options available to speakers. Even so, this principle would suggest that determiners that are more closely related to *said* (such as *this* or *that*) would be more relevant and important competitors than unrelated ones like *a* or even non-determiner alternatives. This principle also ties into the next principle, which discusses how significance is flagged when an unexpected alternative is used.

The final principle of the framework is called The Violations of Expectations Principle (VE). This principle suggests that an utterance is likely to be given special significance within a discourse if or when it violates conversational expectations. This principle, related to Grice's Maxim of Manner, suggests that a response that is unexpected must have a reason for being so, and thus should be paid attention to; hearers are more likely to flag these types of utterances as important to the discourse. If they weren't important, it can and should be assumed that the speaker would not have used them. In other words, this is a principle that says where and how to find or mark

significance, but not what that significance is in any particular instance. This principle most closely relates to SC, the use of which seems to violate conversational expectations, flagging significance. The use of a standard determiner form is, of course, not unusual. The use of *said*, however, is less common, and it seems logical that a speaker would choose to use this form when they intend to draw attention to a particular point or entity within the discourse, or make a particular social move like humor or the assertion of intelligence. This leads to my third research question, which seeks to understand why a speaker might choose to use SC over other viable alternatives, or how a hearer might interpret SC-containing utterances. Chapter 3 of this project uses Acton's framework as the basis for an experimental analysis of the social meaning of SC. In this study, I argue that the unexpectedness of seeing SC, a once formal construction, in an informal setting is an example of Acton's VE principle, and that this unexpectedness flags significance in the discourse. I hypothesize that this misplaced formality communicates a meaning of humor or mock-formalism that is interpreted as funny by hearers.

1.5.2 Further studies in social meaning

A significant portion of this dissertation focuses on the idea of *social meaning*, and how SC carries one that is distinct from any meanings conveyed through other more standard determiner forms. There have been many other studies on how seemingly innocuous speech features can carry robust social meaning. Campbell-Kibler (2009) defines social meaning as "social content tied in the minds of a given speaker/hearer to a particular piece of linguistic behavior". In other words, speakers can

choose features of speech as a conscious act of communicating unspoken meaning, and specific speech acts can cause a hearer to draw conclusions (whether consciously or unconsciously) about a speaker; these conclusions can be broad, such as a regional association, but may also be more specific such as social stereotypes that may accompany a speaker with a certain type or feature of speech. In Campbell-Kibler's study, a single variable (in this case, word final *-ing* vs. *-in*) was manipulated to create stimuli presented to participants, to see if the difference between a word ending in *-in* (as opposed to the same word ending in *-ing*) was enough to elicit a social meaning based interpretation from participants. This study presented participants with matched auditory sentences differing only in the presence of *-in* or *-ing*, followed by a survey task that consisted of questions aimed to understand participants' reactions and feelings toward the speakers they heard. Campbell-Kibler found that speakers who used *-in* were rated by participants as less intelligent than those who used *-ing*, and were also more likely to be considered working-class than middle-class or wealthy. Implementing a similar methodology, Chapter 3 of this dissertation explores the social meaning of *said*, by seeing if matched pairs of written sentences varying only in determiner (*said* vs. *the* or *that*) can receive different reactions from participants. I also use a survey task to ask participants questions about *said*, including open ended questions about what they believe it means and why they believe a speaker would choose to use it.

Social meaning research dates back to the genesis of linguistics as a field; some of the earliest work in sociolinguistics deals with questions related to social meaning. For example, foundational work by Labov (1963) suggests that speakers can make use

of vowel changes such as dipthongization or consonant retention to solidify in-group identity to a particular regional dialect group. Specifically, Labov studies the dialect of English spoken by residents of Martha's Vineyard, an island off the coast of Massachusetts. Long-time residents of the island implore centralization of /ai/ and /au/ diphthongs, which Labov argues functions as a linguistic means to distance themselves from the island's summer onslaught of tourists, solidifying their identity and tie to the island. Labov also asserts that residents of the Vineyard were likely to maintain the pronunciation of word final and pre-consonantal /r/ as a way of distancing themselves from the Boston identity and dialect, which at the time heavily featured an r-less system. This too served as a way of solidifying Martha's Vineyard residents' island identities.

These linguistic observations move beyond variation in a broader sense into the realm of social meaning, where specific features of language can have a distinct meaning and significance within a social group. More recently, studies have explored speech phenomena of all types, from sounds to fillers, to entire words and phrases, and how these sometimes seemingly insignificant speech acts contribute to meaning in discourse. Stubbe & Holmes (1995), for instance, look at the social meaning of "exasperating expressions" such as *you know*, and *eh* in New Zealand English. This study shows that speakers employ these exasperating expressions to show solidarity and shared understanding with their addressees, whereas expressions like *I think* function to do the opposite, drawing attention to gaps in shared understanding between speakers and hearers.

Another study by Norrick (2007) shows that the word class of interjections is an ever growing category with a broad scope of functions including marking contrast, elaboration, and discourse transitions. Still, Bolden (2009), shows that *so*, while commonly used inferentially, is also commonly used as a discourse marker that can be used for the functions of discourse coherence, or to promote a conversational item as furthering the speaker's conversational goals. Bolden uses a corpus of tokens of *so* taken from spoken conversation to construct an analysis about its function in discourse, finding that it is often used at the beginning of a sentence to mark the beginning of an impending action such as a question.

1.6 Conclusion

The *Said Construction* has, to my knowledge, not been previously studied, and this project therefore serves as a starting point for future research on it. The main studies on which this dissertation is based all have roots in discourse pragmatics and/or sociolinguistics. As stated in the research questions above, the main goals of this dissertation are to understand how *said* is used and interpreted within discourse in both Spanish and English. With this in mind, it is clear that a discourse pragmatic framework is the best fit for this project, as it provides the tools for analysis and data interpretation. Sociolinguistic literature is also important, as it provides a basis for applying the larger pragmatic concepts into specific groups of language users--in the case of the studies presented in this dissertation, this group is speakers of American English.

The rest of the dissertation will be laid out as follows: chapter 2 presents a corpus based analysis of SC in order to demonstrate the history of the construction, its evolved usage, and its discourse properties. This chapter serves as a foundation for the rest of the dissertation because it helps motivate the social meaning studies presented in the chapters that follow by showing that *said* functions as a determiner in discourse and may serve a purpose that cannot be accomplished by other determiners. Chapter 3 will present an experiment designed to elicit participants' reactions to and interpretations of *said* in online discourse. This experiment is partially motivated by chapter 2's corpus study, which shows that the usage of *said* has changed over time in form and across genre. Chapter 4 presents results of similar corpus-based and experimental studies on the Dicho construction (DC) in Spanish.

CHAPTER 2

A corpus-based analysis of SC in English

2.1 Introduction

A central idea to the field of Linguistics is that language, by nature, is dynamic. Languages are constantly undergoing change; some changes take centuries, such as the slow regularization of irregular past tense verbs in English. Some changes happen more quickly, as in the instance of nouns becoming verbs; Google, for instance, was merely the name of an internet search engine 20 years ago and is now widely accepted as a verb for doing the act of looking something up on the Google search engine. Some changes are multidimensional, as is the case with the *said* construction (SC), which seems to have changed over time in both form (in the sense that the once widely present determiner *the* is now commonly omitted) and function (from referring adjective to determiner-like). The word *said*, when not used as a verb but instead appearing before a noun, has traditionally been categorized and accepted as an adjective. A closer look at recent usages of this construction, however, shows that it now seems to behave more like a determiner due to its syntactic distribution and relationship to other information in a discourse. Furthermore, *said* can now be used to communicate a unique social meaning, as will be discussed in chapter 3.

The purpose of this chapter is to provide a detailed overview of the *Said Construction* including its past and current usages, aiming to answer the following questions: 1. With what types of information does *said* most frequently occur? 2. How has this construction changed over time? 3. In which genres of discourse is this construction most frequently found?

To carry out this analysis, I use corpus-based methods to observe patterns in genre, information status, and diachronic change in SC. Collecting tokens of SC from the Corpus of Contemporary American English (COCA) and from other naturally occurring sources of discourse such as blogs, social media, television, and books, I show that SC demonstrates diachronic change in syntactic role and genre of use (Davies 2008). While older tokens of SC were relegated to formal genres and were more likely to possess an additional determiner (*the said N*), tokens uttered more recently than 1950 do not as frequently use this additional determiner and are more commonly found in informal genres. I argue that without the preceding determiner, *said* itself displays determiner-like properties and functions as such in the discourses in which it is used. I also show that SC favors a narrow definition of givenness due to its relationship with other information in a discourse. Furthermore, this chapter provides background on SC that is foundational for motivating and understanding the studies presented in concurrent chapters; chapters 3 and 4 focus on the social meaning of SC, which I argue is rooted in the construction's diachronic genre shift.

This chapter will first provide an overview of the history of the construction. From there, I will give a brief review of relevant literature on determiners, givenness, and

information status. I will then present the methods and results of a corpus analysis from the Corpus of Contemporary American English, showing *said*'s information status, how it patterns across genres, and how it has changed over time. Finally, I will use some of the tokens of SC from the corpus analysis to specifically discuss the construction's usage and distribution from an information status perspective.

2.2 History

The *said* construction (SC), a widely used but understudied construction in English, has evolved over time in both grammatical status and usage. Traditionally described as an adjective[3] in dictionaries and grammar guides (Merriam Webster 2018, Macmillan 2019, grammarist 2014), this usage of *said* has historically been associated with formal or legal jargon (Tiersma 1999). Its roots trace back to Imperial age Latin, where forms such as *antedictus*, which translates to *aforesaid* or *aforementioned*, were common (Norberg 1980) (although the origin of the actual word-form *said* is Germanic). These constructions carried into medieval Latin, and were in turn passed onto other European languages; many Latin-derived languages still use constructions similar to SC with similarly translated words in order to convey formalism (Norberg 1980). For example, Spanish uses *dicho* in a similar way to our English SC, which will be discussed in more detail in chapter 4. Example (1) demonstrates this Spanish usage:

(1) Vamos a hablar de Cáceres: **dicha ciudad** fue construida en 25 BC.
We go to talk of Caceres: **said city** was built in 25 BC.
And now we come to Caceres: the/said city was built in 25 BC.

[3] These cited dictionaries have separate entries for adjective *said* and past participle verb *said*.

The earliest tokens of SC date as far back as the 1300s, as evidenced through the Oxford English Dictionary. In the Corpus of Contemporary American English (COCA), the oldest available tokens are from the 1700s. These early COCA tokens have a strong tendency to be preceded by an additional determiner and appear as *the said N*. Newer tokens, on the other hand, do not as frequently have this additional definite determiner, appearing more regularly as bare *said N*. Many of these available older tokens from the 17th-19th centuries are from legal documents such as land contracts, treaties, and other types of agreements between parties. One such example is from the Treaty of Augusta, of 1768, which discusses the fate of the land of the Creek Indians, and says that…

> (2) *"the Lands reserved by **the said Creek Indians**, for their own use, should be distinguished from those Ceded to His Majesty in the Province of Georgia."*

Tokens like (2) are prevalent in the COCA and are mainly taken from academic and historical archives. Towards the latter part of the 20th century, however, corpus tokens begin to demonstrate a shift away from formal genres, and are more commonly found in novels, news, and magazines. Furthermore, these newer tokens are more likely to be missing the additional determiner, appearing only as *said N*. This subtle change in form also seems to be related to the types of genres in which the construction has been used during the past century; newer tokens lacking the additional determiner tend to appear in informal registers such as magazines and spoken conversation, whereas tokens that have maintained the additional determiner are from the more formal genres in the corpus such as written news and academic discourse. In this respect, the book "Language and Law" (Tiersma, 1999) says that while forms like *said*, *aforesaid* and

aforementioned were once widely used in legal jargon, SC is now (at the time of Tiersma's publication) considered obsolete in legal discourse. This is somewhat reflected in the COCA, which shows a drop off in uses of SC in academic genres, while simultaneously showing an increase in registers of discourse more closely associated with the layperson, such as magazines and, more recently, web-based texts.

2.3 Said as a determiner

In order to explain the use of *said* as a determiner, it is important to consider other work on determiners. Much of this past research is discussed in more detail in chapter 1; this section will provide a brief review of key studies involved in this analysis, for the purpose of framing where I believe *said* is situated within the literature and motivating a definiteness and information status-centric analysis. The works I have chosen to reference here were selected in order to better understand the types of information with which various standard determiners are most likely to occur. By understanding the distributions of other determiners in relation to the information preceding them, we can begin to observe and understand how *said* may be different.

Gundel, Hedberg & Zacharski (1993) establish a hierarchy for demonstrating the types of information with which certain common referential expressions such as *a*, *the*, *this*, *that* and *it* are most felicitously used (see chapter 1, figure 1). The authors introduce the idea of cognitive status, which refers to what is in a hearer's mind at the time of utterance. The speaker's beliefs about the hearer's knowledge at the time of utterance informs their choice of determiner; likewise, a speaker's use of a particular

determiner assumes a certain level of knowledge from the hearer. For example, an indefinite determiner *a* is most commonly chosen if the speaker is referring to something with which they believe the hearer is unfamiliar. More familiar entities will be used in conjunction with *the* or *that N*. Other past work has constructed theories of definiteness based on similar ideas: that definite referents must be distinguishable and identifiable from other possible referents, even if they are not specifically known (Hawkins 1978, 1991, Holmback 1984, Abbott 1993 *inter alia*). As we will see in this chapter, *said* is almost always used with information that has been previously mentioned and is not used to refer to an entity present in the situational context of the dialogue.

2.3.1 Information Status

This notion of cognitive status is directly related to other work on information status, which also has to do with what the speaker believes the hearer knows at the moment of an utterance. A broad purpose of information status research is to understand how certain types of entities are most likely to appear in a discourse; this includes things like NPs and how they are used with various types of articles, as well as how they relate to their antecedents. For instance, It has been maintained that information is either new, old (also called *given*), topicalized, or focused; all of these ideas have been addressed in the literature and defined differently by different authors (Haviland and Clark 1974, Li and Thompson 1976, Birner and Ward 1994, inter alia) but are roughly categorized as follows: old information has already been mentioned, or is already shared between speakers by means of common world knowledge, extralinguistic context, or (debatably) inference. New information is introduced into the

discourse for the first time. Topicalized information can be considered the background information to the discourse, whereas focused information is what is currently being discussed or emphasized in the discourse.

This study in particular focuses on old vs. new, categories which can be divided even further into what is old/new to either the hearer or the discourse itself (Prince 1992). While it is generally understood that new information is preceded by *a* and old with *the*, many exceptions exist, leading to two related problems: there is not a theory of definiteness that fully accounts for the distribution of *the*, nor is there a theory of givenness that is conventionally agreed upon. A strictly linguistic model of givenness (such as presented in Schwarzschild 1999) considers given only that which as been explicitly mentioned in the discourse. A model of givenness that also considers hearer status (Prince 1992) would include that which has been mentioned, in addition to that which the speaker believes the hearer knows based on the immediate extralinguistic context of the conversation (such as entities in the physical setting) as well as common world knowledge (such as “the stars” or “the government”). Finally, a broad model of givenness (as in Birner 2006, for example) would take all of the above into account, while also including that which can be inferred, such as *a show* → *the audience* (Birner 2006, example 6).

While this dissertation does not claim to solve either of the definiteness and givenness problems mentioned above, I believe adding *said* to the equation can contribute new and valuable information to these discussions; a large portion of this corpus study focuses on looking at the types of information that precede SC, paying

particular attention to whether *said* is most likely to appear with a noun that has been previously mentioned, or is inferred from what has been mentioned. Further information pertinent to this study can be found in chapter 1.

2.4 Corpus Analysis: Methods

This section presents the results of a corpus based analysis on the genre distribution, information status, and historical use of the Said Construction. The Corpus of Contemporary American English was used for this analysis, selected based largely on the fact that it is a balanced corpus; each genre in the corpus represents an equal number of words and the words added each year are equally split among these 5 genres: spoken language (including sources like NPR shows and 60 minutes), academic writing, fiction, magazines, and news. The corpus adds 20 million words of data each year, and is currently about 560 million words large. The COCA's algorithm allows for searching by word, part of speech and lemma (such as searching the verb "break" and getting all forms, such as "broke", "broken", etc.). It provides frequency data and genre distributions of search results, making it an ideal source for the data needed for this study.

To obtain tokens for the present analysis, the COCA was searched for instances of adjectival *said*, followed by a noun or an adjective, using the search term "said_j* NOUN". The "j" is used in the COCA to denote adjectives, so "said_j" means "said used as an adjective." At the time of initial collection (August, 2018) this search resulted in 1013 total tokens. Additionally, the corpus was searched for adjectival *said* followed by

an adjective (using “said_j* ADJ”) in order to account for tokens of SC that included an intervening adjective between *said* and the following noun. This search resulted in only 91 tokens at time of collection. Tokens were read individually to ensure the proper usage of *said*, which was determined based on location (usually directly after a verb) as well as the presence of a different verb in the sentence, signaling that *said* was not being used as a main verb or past participle. This hand coding was necessary since a majority of resulting tokens from the search had *said* coded incorrectly as an adjective but used as a verb, as in (3).

> (3) Feminists were divided; some **said women** should be free to serve as surrogates and get paid for it if they so chose. [Mother Jones Magazine, Vol. 36 Iss 1. From COCA]

Tokens such as (3) were excluded from the sample, as were tokens in which an antecedent was not visible in the limited context provided by the COCA, and could not be found via other search platforms such as Google or JSTOR. Of the 1013 original search results, only 315 were usable once exclusions were applied[4].

Default genres assigned by the corpus were saved: magazine (MAG), novels (FIC), spoken language (SPO), news (NEWS), and academic (ACAD). Where necessary, however, the year of writing/utterance was adjusted for accuracy. For instance, many examples of SC were from academic sources from the 20th century, but were directly quoting land treaties and contracts from the 18th century. In these cases, the year was adjusted to reflect the year the treaty or contract was written and the token

[4] It should also be noted that there were potentially instances of SC that were missed due to *said* NOT being tagged as an adjective, as in instances of *said* tagged as a verb but actually used in the *said* construction. *Said* in its verb usage was never searched or used in this analysis.

of SC was originally used. For the sake of this study, no tokens from 2018 were included, as at the time of analysis 2018 was not yet finished and new data had not been added to the corpus. Furthermore, I separated out the 54 tokens uttered/written before 1950 to be analyzed on their own for a portion of the analysis, in order to take a closer, more in depth look at how the construction has changed specifically in the last 30-40 years, directly before and after the introduction of the internet into the mainstream and the genesis of web-based discourse.

In addition to the tokens acquired from the COCA, part of this analysis uses 90 additional tokens of SC which were collected from other naturally occurring sources of discourse, largely social media and websites/blogs. To account for these genres, two additional genre categories were created: SOC and WEB. These two additional genre categories were kept separate from each other due to the different uses of language within them; social media postings are generally much more succinct than speech found in blogs and websites. Twitter has until recently had a 140 character limit, and social media interactions are often kept to a sentence or two. Blogs and other types of web-based discourses, on the other hand, can be much more lengthy and verbose than social media. It is also important to note that these two categories were only used for the information status portion of this analysis; SOC and WEB tokens were kept separate from the COCA tokens for the sake of balance and accuracy in analyzing the COCA data for diachronic change and syntactic function.

405 total tokens (315 COCA, 90 “in the wild”) were organized in a spreadsheet and were coded by various additional criteria besides year and genre; tokens were

assigned a 1 or a 0 for presence or absence of an additional definite determiner (***the*** *said N*), 1 or 0 for the presence or absence of an intervening adjective between *said* and the noun (*said* ***blue*** *car*), and a numbered category to account for the *said N*'s discourse relationship to its antecedent. These categories are broken down and described in more detail in Table 2.1.

Cat.	Information Type	Token Example
1	Explicitly mentioned	The system is the central problem, but **the "tool"** is definitely complicit, especially if **said "tool"** is... a sentient being.
2	Subtype/type or property/missing property	The final sentence reads: '**Vietnam**, hot damn.' Since I was soon to ship out for **said country** myself, I remember it very well.
3	Synonym	Just imagine you're moving tomorrow, what all would you actually need or want to take along? Odds are, it wouldn't be much. We hold onto **so many things** because of what we attach to **said items**, whether it be guilt, buyer's remorse, or something along the lines of "I'll use that at some point!" but really, we don't need excess to live exceptional lives.
4	Inference	Schuerholz isn't prone to **hyperbole**, but he appeared guilty of **said offense** after the December trade for Tim Hudson.

Table 2.1: coding categories for SC data

Category 1 marked tokens of SC in which the noun was exactly the same as its antecedent. Category 2 denoted tokens where the antecedent was a subset or property of the noun mentioned in the SC token. In the example provided in table 2.1, “Vietnam” is a (singleton) subset of “country”. Other examples in this category included properties like “10 lbs” in “a 10 lb baby” which were then absent in the SC, as in “said baby”. In these types of examples, one could argue that a 10 pound baby is a type of baby; this

type of use of SC was the most common in this category. This category also included tokens where the antecedent was multiple words or a list that were then summarized with a single word in the SC (again, a subset/set type of relationship), such as *silly putty, coloring books, crayons, and other little odds and ends → said items*. Category 3 was similar to category 2, but the subtle difference was that the antecedent and noun were genuine synonyms of one another. In table 2.1, we see tokens such as *things → said items*, and other data include examples such as *someone → said person*, and *puke → said vomit*. The final category was reserved for inferences. The example shows *hyperbole → said offense*, which requires the reader to draw an inference that a hyperbole is indeed an offense. For the sake of clarity of categorization, another example from category 4 would be *sheets that felt as if they exceeded 300 thread count → said silky sheets*, building the inference that sheets with a higher thread count would be silky. While I acknowledge that this categorization system is not perfect, I can confidently say that all collected tokens of SC fit clearly into one of these categories, with very few exceptions that could be argued to fit into two categories. For these problematic cases, I asked for feedback from multiple colleagues, eventually categorizing the tokens into the category with majority consensus.

2.5 Results

2.5.1 Information Status

Totals for each of these categories were as follows:

Category 1	n=203 (50%)

Category 2	n=160 (40%)
Category 3	n=29 (7%)
Category 4	n=13 (3%)

Table 2.2: I.S. Results, English

Category 1 (explicitly mentioned) was the most common, with 203 tokens. Category 2 (subtype/type) was the second most common, with 160. Categories 3 (synonym) and 4 (inference) were the most uncommon, with 29 and 13 tokens, respectively. From these category totals, we see that *said* has a clear preference for the types of information with which it can occur. Out of 405 tokens, 363 (90%) of them appear with information that has been mentioned, either explicitly or via a subtype/type/property relationship (categories 1 and 2), and only 3% of tokens were used with inferred information. Although theories of givenness vary in breadth, even very restricted theories consider the types of information in categories 1 and 2 given (see section 2.3.1 for information about these varying models of givenness). An exact binomial test reveals that the likelihood of a token using *said* with an inference is no more common than chance, at a level of [p=.02]. Furthermore, this restricted use of *said* with inferred information can perhaps provide support for the preference of a linguistically based model of givenness that does not include varieties of inference and/or world knowledge.

2.5.2 Genre of Use

While these totals are helpful in providing clues about the types of information in which *said* is most commonly used, it leaves open questions of genre usage and how the construction has changed diachronically. This portion of the study used only tokens

found in the COCA, excluding “in the wild” tokens for the purpose of retaining a balanced sample. Using R, relationships between variables were calculated to show correlations between tokens of SC across year and genre, as well as the presence or absence or an additional determiner. A general linear model was used to examine the relationship between the absence of an additional determiner within the genre categories listed above. Results showed a significant relationship between the lack of an additional determiner and genres of fiction, magazine, news, and spoken data, as shown below in table 2.3.

	Estimate	Std. Error	t value	pr(>|t|)
Intercept (det)	0.65217	0.0351	18.60	<2e-16
Fiction	-0.5558	0.0542	-10.26	<2e-16
Magazine	-0.5252	0.0589	-8.91	<2e-16
News	-0.6228	0.0734	-8.49	9.00e-16
Spoken	-0.5022	0.0911	-5.51	7.43e-08

Table 2.3: GLM data, additional determiner as intercept, within genre categories

Figure 2.1 below also shows the usage of *said* across genres, over the full timespan of tokens used from the COCA sample. Of the 56 tokens of SC older than 1950, only two of them were not in the academic genre.

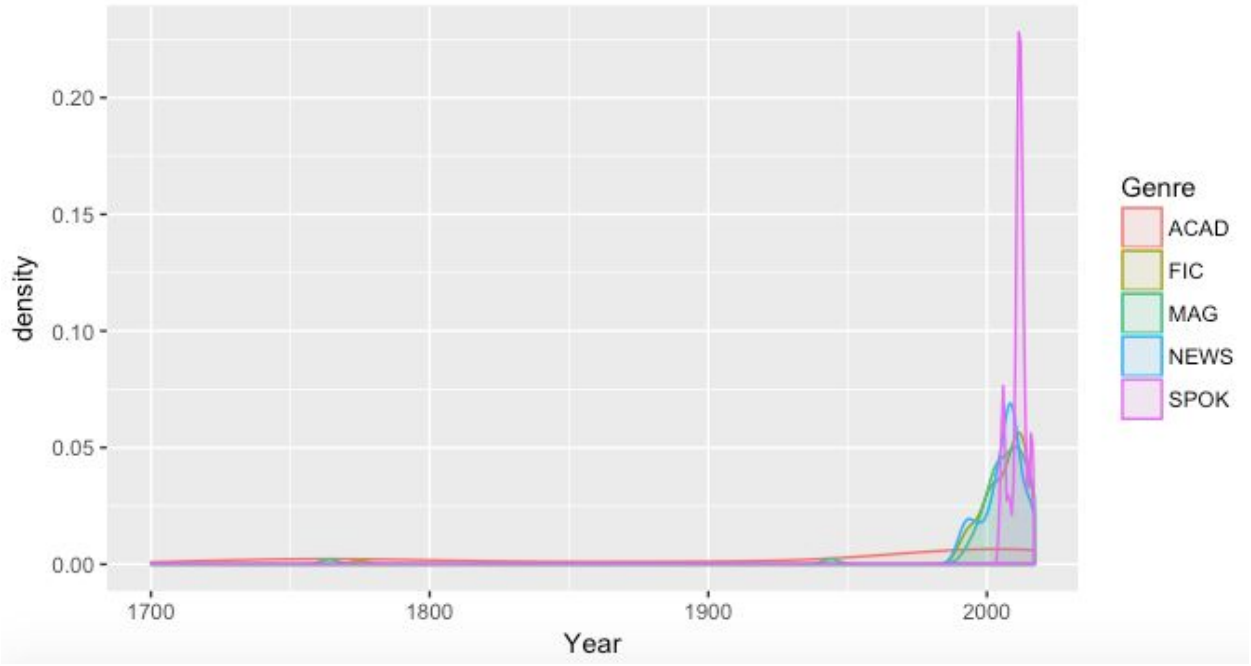

Figure 2.1: Genre distribution timespan

Zooming in on the most recent 40 years of data, we see a decline in academic (mostly legal) discourse while simultaneously an increase in more informal genres such as spoken language. This pattern potentially supports the aforementioned notion from Tiersma 1999 that SC is now less frequently used and potentially even considered obsolete in legal/formal discourse.

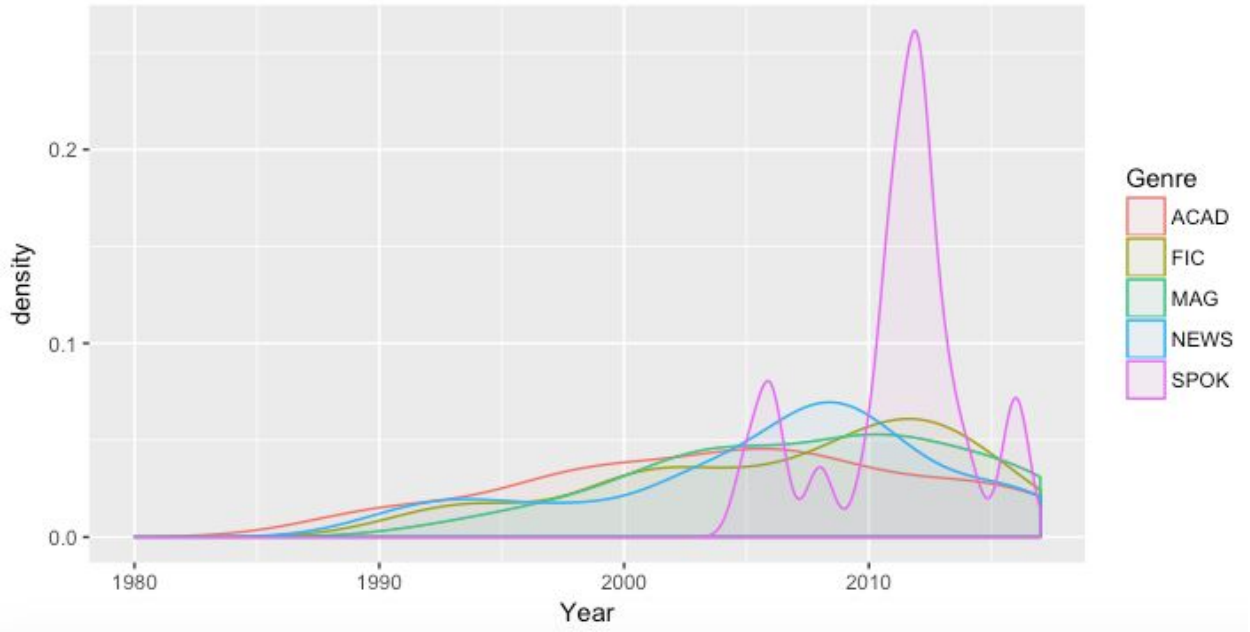

Figure 2.2: COCA tokens across genre, 1980-present

2.5.3 Diachronic Change

I also calculated how the syntactic form of SC has changed over time. From the corpus data it seems that earlier tokens of SC in the corpus are more likely than recent ones to demonstrate the presence of an additional determiner before *said*. To test this, I used a general linear model to test the relationship between the year of utterance of the token and the presence or absence of an additional determiner. For this portion of the analysis, I excluded "in the wild" tokens in the SOC and WEB categories to ensure that the sample included only balanced data from the COCA. Results of this experiment show a significant effect between year and presence of an additional determiner, with newer tokens being significantly more likely to lack a determiner [$p<.001$], as shown in table 2.4. This is also displayed in figure 2.2, in which I show patterns in how SC has

been used both with and without an additional definite determiner over the past few decades. This shows that over time, SC has become less likely to be used with an additional determiner; the most recent tokens of SC are more likely to stand alone as *said N* while tokens older than approximately 15 years old are more likely to take the form of *the said N*.

	Estimate	Std. Error	t value	pr(>|t|)
Intercept (det)	5.0323	0.383	13.14	<2e-16
Year	-0.002	.00019	-12.38	<2e-16

Table 2.4: Fixed effects for English COCA data, determiner as factor

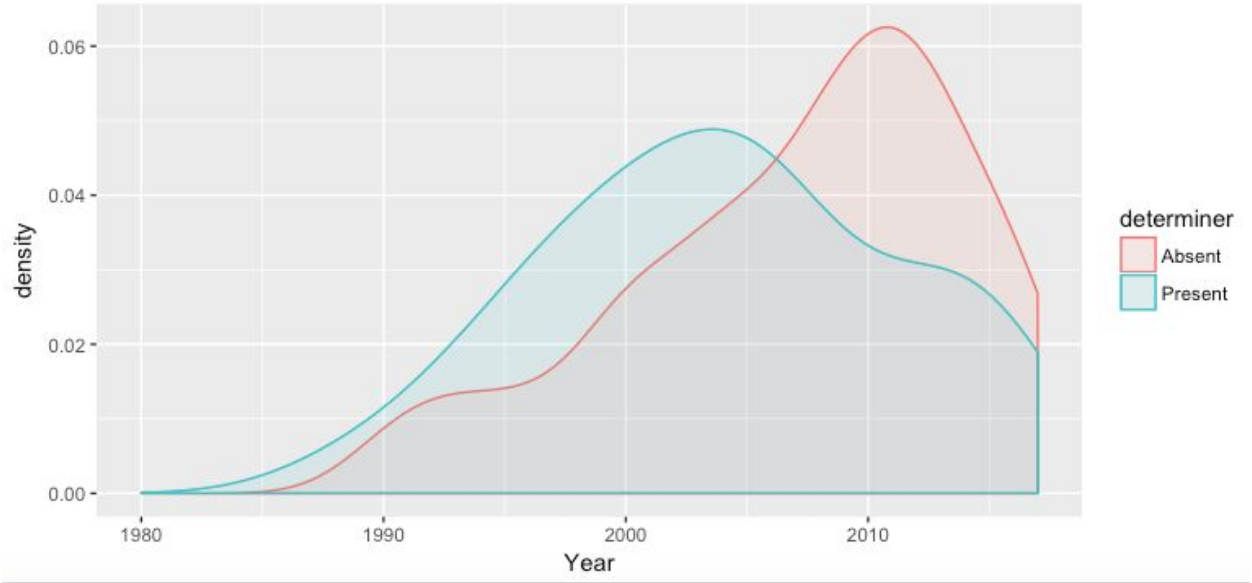

Figure 2.3: COCA tokens, presence or absence of additional determiner preceding SC

The above analysis only includes default genres from the corpus, excluding web based tokens. I acknowledge a potential problem with this analysis, which was how to account for these two added genres (SOC and WEB) in light of the fact that the COCA is a balanced corpus. Since I did not take a COCA sized sample of social media and web based discourses, I did not include data from these categories in the statistical analysis above or in the graphs shown in figures 2.1 and 2.2. I can, however, attest that BYU's new iWeb corpus which featuring web based discourse from only 2017 has over 35,000 tokens of SC, providing support for my argument that this construction has dramatically increased in popularity over recent years and specifically in web-based discourse. Considering that iWeb is 25 times the size of the COCA, Simplifying the number iWeb of tokens to match the frequency of the COCA results in around 1400 tokens which is not much more than the 1013 from the original COCA search, it is important to remember that iWeb includes only one year of data whereas the COCA spans several centuries. In other words, there are more tokens of SC in one year of iWeb data than in the entire timespan of the COCA.

2.6 Additional Corpora

As stated above, an undeniable benefit of using the COCA for this type of analysis is the fact that it is balanced for word count across the included genres. The BYU collection of corpora, where COCA is housed, also includes several other corpora that are insightful in this analysis. I collected additional tokens of SC from some of these corpora as a way to expand my search into a wider range of genres and discourse

types. Although these tokens were not included in the statistical analysis presented above, the searches proved insightful and are worth briefly mentioning.

2.6.1 iWeb

The iWeb corpus was released in 2018, towards the end of the analysis process for this project, and contains data only from 2017. This corpus consists of 14 billion words, all from language on the internet from primarily English speaking countries such as the US, Canada, the UK, Australia and New Zealand. I did not include data from this corpus in the main part of my analysis for a few reasons. First, since iWeb includes only internet data, it didn't provide as wide a scope for this project as COCA which is balanced across the several genres mentioned above. Additionally, it only contains data from 2017[5], so it would not be useful in looking at diachronic change. Finally, this corpus does not have an option to search only websites from the US, and since this portion of the project focuses only on American English, I did not want to include data from other English speaking countries.

As mentioned above, I did conduct a search using the same search terms in order to see if there were any patterns of SC that would emerge looking only at recent internet data. Using the corpus' maximum search of 1000 contexts, this query yielded over 35,000 tokens of potential SC, supporting the notion that this construction has dramatically increased in popularity over recent years and specifically in web-based discourse. In other words, 200+ years of COCA data produced only a fraction of the number of tokens of SC than that of a single year of web data. And although the iWeb

[5] There were, however, some tokens (much like in the COCA) that used SC in quotation of older documents.

corpus is 25 times the size of COCA, even 1/25 of the results from my iWeb search is still more tokens (in one year) than that of the entire timespan of the COCA[6]. As mentioned, however, these patterns do not speak to American English specifically, as the corpus contains data from several English speaking countries. Even so, it is interesting and relevant to observe the prevalence of SC in this corpus, motivating future research focusing on the usage of SC in web-based discourse.

2.6.2 SCOTUS

I consulted the Corpus of US Supreme Court Opinions, which will be referred to as the SCOTUS (as established by the BYU corpora team) in order to address the claim made by Tiersma (1999) that SC was a once widely used legal construction that had since become obsolete. I also wanted to further address the question of whether or not SC was once more widely used in formal language than informal, and if it has undergone a change even within the more formal genres where it has traditionally been found. The SCOTUS is a 130 million word corpus of American Supreme Court decisions from the 1790s to the present. This corpus is not balanced across years; there are over 14 million words from the 1980s, for instance, and only 9 million for the 1990s. The exact breakdown of words and texts available in this corpus per decade is available in Appendix A.

This corpus was used in order to get a clearer picture of how SC may be used in legal language, and to see if this usage has changed over time. As mentioned above,

[6] Furthermore, a random sample of 200 iWeb tokens had only 10 instances of incorrect tagging of *said* being used as a verb, so a further analysis from this corpus would likely result in fewer tokens needing to be thrown out due to the wrong form of *said*.

Tiersma (1999) asserts that this construction is considered somewhat archaic and obsolete in modern legal language. In order to best compare the data from this corpus with that collected in COCA, I analyzed it from two different angles; I looked at data from each corpus across all included time periods, and also from 1950-present. This allowed for both a holistic picture of how the construction has been used historically, and also how it is being used more recently, and if this has changed. I used the same search terms as specified above for the main analysis portion of this chapter in order to obtain tokens of *said* tagged as an adjective, followed by a noun: said_j* NOUN. This search resulted in over 122,000 tokens of potential SC, a surprisingly large number considering this corpus is about a quarter of the size of COCA. A random sample of 200 was taken from these results and coded for year, the presence or absence of an additional determiner, and the presence or absence of an intervening adjective, as in the main analysis above. This random sample did not contain any tokens that were more recent than 1972[7]; all tokens ranged from 1787-1972. Furthermore, when the sample was ordered by year, besides the most recent six tokens, all tokens were older than 1950. 120 of 200 tokens (60%) lacked the additional determiner, and only two tokens had an intervening adjective. The COCA data included 315 tokens ranging from 1500-2017. Of these, 219 (70%) lacked the additional determiner. It should be noted that the vast majority (97%) of these 219 tokens are newer than 1950; only 7 pre-1950 tokens of SC from the COCA lacked an additional determiner. These data seem to support the notion that SC was once more common in legal language than it is now. Furthermore, the fact

[7] One tokens appeared in the random sample that was from 2008, however this example was one which was included in a direct quotation from an older document, dated 1787.

that the SCOTUS data was more likely to contain the additional determiner also speaks to the pattern that emerged in the COCA data, that SC in its original form is more likely to appear in genres that can be considered “formal” such as legal and academic discourse.

I also conducted a search looking only at 1950-present, as these are the years from which the majority of tokens from the COCA came. From this narrowed down time frame, the SCOTUS contained only 2972 tokens of SC, or only about 2% of the total tokens of SC in the corpus. Of these, only 700 contained an additional determiner; 76% of these tokens were lacking a determiner, in contrast with the 61% of tokens over all time (from the smaller sample size) that were missing a determiner. This trajectory of newer tokens lacking an additional determiner is also seen in the COCA data: 212 of 261 post-1950 tokens (81%) lack a determiner. These patterns show that the overall direction in which SC is heading is one without a determiner, and that this is the case in formal and informal genres. This change, however, does seem to be happening more quickly in more mainstream and informal genres of discourse.

2.7 Conclusion

2.7.1 Said as a determiner

While the analysis presented in this chapter speaks to the idea that *said* behaves like a determiner in the way it interacts with previous information in the discourse, there are, of course, counter arguments for considering *said* in this way. Who’s to say *said* is a determiner? What if, in cases of *said* N, there is a null determiner and *said* is still an

adjective? These are valid questions and I do not necessarily intend to say that *said* is not an adjective. I would like to propose, however, that it *behaves* very similarly to other determiners in the way it relates to other information in a discourse; for instance, like *the*, it is sensitive to the conversational status of its antecedent, and is most likely to be used with information that has already been mentioned. And if it behaves like a determiner, what else can we gain from analyzing it as such? Operating under a linguistically-based definition of givenness (one excluding inferences, extralinguistic conversational context, and common world knowledge), we can see that *said* is most commonly used with given information. Consider the following examples, taken from the COCA:

> (4a) For the past seven years, Darrell and I had been living in New York City, where "harvesting" a Christmas tree involved forking over an exorbitant sum for **a scrawny sapling**, then lugging **said sapling** up five flights of stairs to our itsy-bitsy apartment. [Country Living Title: An Evergreen Tradition. Jan. 2012]

In this example, *said sapling* is given, licensed by the explicitly mentioned antecedent *a scrawny sapling*. According to the coding convention I proposed above, this example is type 2, as the antecedent and SC NP form a subtype/type relationship. This use of SC would not work if information were ordered the other way around. Consider 4b:

> #(4b) For the past seven years, Darrell and I had been living in New York City, where "harvesting" a Christmas tree involved forking over an exorbitant sum for **a sapling**, then lugging **said scrawny sapling** up five flights of stairs to our itsy-bitsy apartment. [adapted from Country Living Title: An Evergreen Tradition. Jan. 2012]

This type of usage is not attested in the sample collected from the corpus, and it is not difficult to see why. Not all saplings are scrawny saplings, and the mention of *a sapling* does not make salient any descriptive attributes. Tokens like the example in 4a, on the other hand, are widely attested and used; all scrawny saplings are saplings, and mention of *a scrawny sapling* is enough to establish *sapling* as given information for the remainder of the discourse. Using this same SC token shown in 4a and b, we can construct another example using standard determiners in place of *said*. In the case of (4c), either alternative determiner would be acceptable, as would a pronoun such as *it*.

> (4c) For the past seven years, Darrell and I had been living in New York City, where "harvesting" a Christmas tree involved forking over an exorbitant sum for **a scrawny sapling**, then lugging **the/that/said sapling** up five flights of stairs to our itsy-bitsy apartment. [adapted from Country Living Title: An Evergreen Tradition. Jan. 2012]

Another characteristic of *said* that makes it seem determiner-like and sets it apart from other determiners is that it seems to lack the ability to occur in generic NPs. As mentioned in chapter 1, a definite determiner or demonstrative can be used to refer to either a generic or specific entity:

> (5a) *The dog* has been a loyal companion to mankind for centuries.
> (5b) *The dog* got out of the backyard last night.

(5a) represents this generic usage, referring to dogs in general. 5b refers to a specific dog, presumably one belonging to the speaker. In both of these examples, one can imagine the utterance occuring at the beginning of a discourse; either could be used and understood without prior context. *Said*, on the other hand, could not be used in a sentence like (5a), even with an established referent. Consider (5c) and (5d):

#(5c) Many early depictions of human history contain references to animals, such as horses, mammoths, dogs, and cats. In fact, **said dog** has been a loyal companion to mankind for centuries.

(5d) Bailey is in trouble today. **Said dog** got out of the backyard last night.

The generic reading of *said* in (5c) is not attested in the corpus sample and does not seem to be felicitous. The specific reading in (5d), however, is common and acceptable. There do seem to be potential cases in which *said* can be used in conjunction with a generic NP introduced by a definite determiner. Uses like this were not attested in the corpus sample, but it is reasonable to say that they may exist and be used. In these cases, the use of *said* creates a shift from generic antecedent to more specific entity in SC[8]. Following the pattern we see in the corpus, a referent must be established before use of *said*, and is most felicitous if a subtype/type relationship is formed:

?(6) The cockatoo, native to Australia, is known for its crested feathers. **Said bird** can also make a great pet.

In this example, the generic use of *cockatoo* establishes *bird*, making the use of a *said* DP felicitous, as its referent is at that point Hearer-old. It is still being used somewhat generically since it is not referring to a specific bird (such as someone's pet), but with an understanding that the intended reading is referring to a type of bird; "Said type/species of bird can also make a great pet."

As mentioned above, *said* has traditionally been considered an adjective and is labeled as such in dictionaries and grammar guides, as well as in corpora as shown above. Most adjectives, however, are generally not sensitive to information status; they

[8] See chapter 1 for a more detailed explanation of taxonomy and ordinance, a la Rosch 1976

can co-occur with old and new information, are used with definite or indefinite determiners, and do not take into account the hearer's knowledge of the topic of discourse. These are some of the most significant properties of *said*, and the behaviors that set it apart from adjectives and make it seem more determiner-like in function.

2.7.2 Summary

Through the study presented in this chapter, I have provided a summary of the distribution, information status, diachronic change, and genre distribution of the *said construction*. Broadly, the analyses presented here provide evidence that this construction has implications on existing theories of definiteness and givenness, as it interacts with these notions differently than other determiners; it favors a strictly linguistic model of givenness, and is used with definite information with a more strict distribution than other standard determiners and demonstratives. When analyzed as a determiner, SC provides a new angle from which to look at these existing theories, and may provide some of the information necessary to confront existing holes in these theories. For instance, as mentioned in chapter 1, there is not a theory of definiteness that fully accounts for the distribution of the definite article in English, nor is there a conventionally agreed upon definition of givenness. I believe these gaps exist because there are numerous counterexamples and exceptions for how every type of sentence and determiner construction (such as existential *there* sentences, for example) interacts with these theories, several of which I have mentioned. But *said* seems to pattern differently enough from other determiners that it may help provide insight that could potentially narrow down the scope of these theories and bring the field closer to a

conventionally agreed upon definition. If the goal, for example, is a theory of givenness that fully (or as fully as possible) accounts for the informational distribution of all definite determiners, a theory that only considers *the* and perhaps *that* would be insufficient; a theory that includes the types of information used in conjunction with *said* is needed. There are at least two potential ways this could be done: either by excluding inferred information from definitions of givenness altogether, or by dealing very intentionally with the contexts in which inferred information can be considered given, and making this a key part of a theory of givenness.

Returning to a summary of this chapter, the corpus-based analysis I have presented highlights that a change has taken and is currently taking place in the usage of this construction. Older tokens are more likely to appear with an additional definite determiner, while newer ones tend to lack this determiner. In the absence of an additional determiner, we can see that *said* functions to refer to information in a similar way, taking on a determiner-like role. The question then shifts away from the *what* to the *why*; in any given token of SC, we can imagine another determiner or demonstrative being just as grammatical, so why would a speaker chooses to use *said* if they don't need to? The following chapter pursues this question, showing that it may function to communicate a specific social meaning that cannot be conveyed through the use of a standard determiner.

CHAPTER 3

Social meaning and the Said Construction

3.1 Introduction

The study on the Said Construction presented in this chapter is rooted in the sociolinguistic concept of social meaning. The field of sociolinguistics is dedicated to exploring questions related to the intersection of linguistic and social performance. Across discourse communities, we find variation in every aspect of language, from speech sounds, to lexical items and entire words and phrases. Variation is found within regional dialects, but also on a much smaller level; there can be distinct patterns of variation even within different social groups in the same community, and even within a single individual.

It is well known that language plays an important role in the construction of group identities, both in defining the social parameters of an individual group and distancing that group from others. For instance, works by Eckert (1989, 2000) and Mendoza-Denton (2008) have investigated the use of language as a marker of identity within high school social groups. In these examples, linguistic performance had social implications beyond functional use; language served as a marker of social status and group identity. Recently, more work has emerged that centers around the idea of *social meaning*, defined by Campbell-Kibler (2009) as “social content tied in the minds of a

given speaker/hearer to a particular piece of linguistic behavior". In other words, speakers can choose features of speech as an act of identity construction, or even to convey a certain persona or mood in a discourse. Furthermore, specific speech acts can cause a hearer to draw conclusions (whether consciously or subconsciously) about the identity of a speaker; these conclusions can be broad, such as a regional association, but may also be more specific such as social stereotypes that may accompany a speaker with a certain type of speech or displaying a particular speech feature. It is this type of meaning that the present project focuses on by looking at the broader discourse community of internet users. I will show that SC contributes a subtle social meaning to the discourses in which it is used and is chosen by speakers to accomplish this purpose instead of a different, more standard determiner form such as *the* or *that*.

As discussed in chapter 1, Eckert (2012) notes that some of the earliest sociolinguistic work done by Labov (1963) touches on issues of social meaning; in his interviews with residents of Martha's Vineyard, Labov suggests that speakers of that particular dialect of English employ use of high centralization in the diphthongs /ai/ and /au/ as an act of resistance against the island's summer onslaught of tourists. Labov also asserts that residents of the Vineyard maintained the pronunciation of word-final and pre-consonantal /r/ as a way of distancing themselves from the Boston identity and dialect which uses an r-less variety of English, and as a way of solidifying their identities as long-time residents of the island. In both cases, these linguistic observations move beyond variation in the broader sense of linguistic differences between regions into the

realm of social meaning, where specific features of language can have a distinct meaning and significance within a social group.

Other sociolinguist work has observed similar phenomena. Trudgill (1983, 150-154) examines the diachronic usage of post-vowel /ɹ/[9] by British rock band The Beatles and shows that in their earlier recordings, The Beatles were more likely to pronounce /ɹ/. This tendency, however, decreases sharply over the years that followed. Trudgill argues that this could have a few different motivations, all based in social meaning. One possible explanation is their gradual shift in musical genres from rock to a less clearly defined musical category. Early pronunciation of /ɹ/ possibly served to construct an identity as a rock band among other bands in the same genre, while their later shift into more poetic, self-written, British-themed pieces served to create distance from the American rock genre. Furthermore, these later songs by The Beatles contained more British sounding rhyming patterns such as *bought* and *short*, which were pronounced r-lessly (Trudgill 1983, 153). Another proposed explanation for this change in /ɹ/ pronunciation is that in their early albums, it is possible that The Beatles were trying to gain popularity and establish credibility with their American audience by pronouncing their words in a more American English-sounding manner. As time went on, however, The Beatles began using a more British sounding, r-less variety of English to solidify their identity as a British band. All of these potential explanations for this shift in /ɹ/ usage are rooted in the idea of social meaning by implying that the linguistic choices speakers make are significant in how they construct their social identities and

[9] I use /ɹ/ in this paragraph to mean any pronunciation of an r-type sound. In the Beatles music, this can be [ɹ], [ɾ], or occasionally [r].

are perceived by addressees[10] in a discourse.

This chapter builds upon the corpus study presented in chapter 2 by further exploring the idea that *said* as a determiner has a different function in discourse than standard determiners, and provides experimental evidence that this meaning may be a social one. The present research applies the social concepts mentioned above to SC, which, though commonly used in English, has not been studied from a linguistic angle. As mentioned in previous chapters, *said* used in this way has traditionally been defined as an adjective in dictionaries and grammar guides, but a closer look shows that this usage of *said* actually displays determiner-like properties in the absence of an additional determiner.[11] As shown in chapter 2, most tokens of SC uttered/written before the middle of the 20th century contain an additional determiner, such as *the said N*, but more recent tokens have a strong tendency to lack this additional determiner which seems to allow *said* to take on a determiner-like role. It is important to note that I am not claiming that *said* is necessarily a genuine determiner. Rather, I suggest that in the absence of an additional determiner, *said* may be able to function as a determiner and that by doing so, it can and should be analyzed as part of the English determiner system. This is evidenced by the lack of another determiner in sentences in which SC is used, and by the fact that *said* is sensitive to notions of information status and givenness in a way that closely mimics other determiners. In this determiner-like role,

[10] I use the term "addressee" to mean any audience for a particular speech act. In other words, this could mean an individual, a small group of people (such as a social/friend group) or even a larger societal group such as in the case of the "speaker" being a band or an actor and the "addressee" being anyone who listens to or watches that form of media.

[11]

said complements the use of *the* by fulfilling the function of definiteness, but also contributes a potential social meaning that I elaborate upon further below. This could potentially be explained by syntactic movement, illustrated in the following trees:

1a 1b

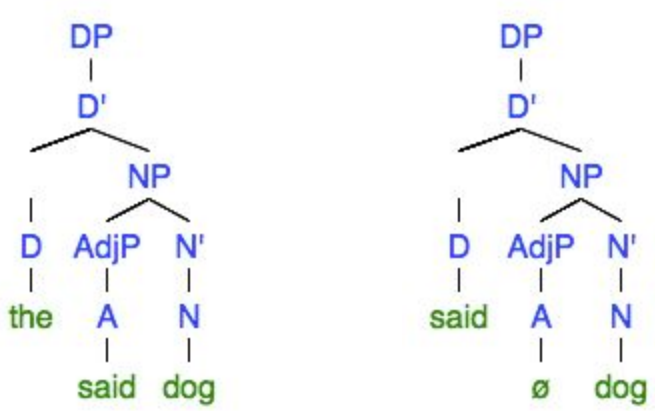

Figure 3.1: Syntax of *the said N* and *said N*

In this type of usage, *said* can fit syntactically in determiner position due to its placement directly before a noun and often after a verb. It can also (though far less frequently, as attested by the COCA data and shown in chapter 2) appear before an adjective, as in *said brown dog*. As introduced above, I argue that *said* used as a determiner can hold unique significance within a discourse; if a speaker is going to choose to use this nonstandard determiner form instead of a standard form like *the* or *that*, there must be a reason for doing so, and I believe that this reason is directly related to social meaning.

While determiners may at first seem an unlikely class of words to convey robust

social meaning, existing studies have shown that this is not the case; they can be solid contributors of meanings within discourse. Foundationally, Lakoff (1974) argues that demonstrative determiners such as *that*, *this* and *those* can be employed by a speaker in order to establish closeness or camaraderie with other discourse participants. Acton and Potts (2014) build on and expand this analysis, showing that the use of demonstrative *that* serves to essentially "level the playing field" between speaker and hearer, showing experimental evidence that speakers use *that* to communicate empathy and shared perspectives.

Acton (2014, 2019) looks at the social meaning of determiners, arguing that determiners can, depending on context, convey distinct social meaning. For example, Acton shows that by using *the* within a group-denoting NP instead of a plural noun (*the Raiders fans* as opposed to *Raiders fans*), the speaker typically situates himself or herself as a non-member of the group to which he or she is referring. This is particularly seen in settings in which group membership is especially salient or important, such as in the political sphere. Acton develops a framework within which to analyze these types of phenomena, namely questions relating to social meaning and variation at the word level. Bringing together meaning-based research across several subfields (specifically semantics, pragmatics, and sociolinguistics) Acton's framework addresses these types of questions. The main principles of this framework, as explained in chapter 1, will be used in the present analysis of the social meaning of SC, in which I argue that SC carries a social meaning of humor and intelligence and that this meaning is distinct from any potential meanings conveyed by other English determiners.

The experiment presented in the following section will address the question of what hearers believe speakers are trying to communicate when using SC. When a variety of more standard alternatives are available, there must be a reason why SC is chosen instead; while chapter 2 showed that *said* does have a relationship to information in a discourse that is unique from that of any other determiner, there is always another determiner that can be used in place of *said* in any given conversational context. Therefore if there are viable alternatives to SC that are more standard in form, and assuming that *said* and other determiners are not in free variation with each other (as shown in Chaper 2), a speaker must have a specific intention in choosing SC.

The broader purposes of this study are to test the hypothesis that SC carries a social meaning, and to try to pinpoint what specific meanings listeners interpret when they encounter this construction. When considering SC's origin in formal and legal language coupled with its shift into standard, everyday speech, we have reason to predict that this contrast between its formal past usage and its informal current usage is meaningful. This leads to the hypothesis that a formal construction in an informal context could lead to an interpretation of humor in the discourse due to its unexpectedness; misplaced formality may have the potential to seem funny to a hearer, as it could be argued to violate Grice's Maxim of Relation (Grice 1975, Goatly 2012, Attardo 2017). In other words, I predict that listeners may interpret SC as funny or humorous when they encounter it in everyday, informal language. It should be noted that I acknowledge a distinction between perception and intention. This study focuses primarily on how discourse participants *perceive* SC and what they believe a speaker's

intention is when using it. I did not pursue answers to questions specifically relating to the *intention* of speakers when they use it. In other words, this study focuses on the addressee and not the speaker. It is likely (and worth the time spent on further research) that speakers' intentions and addressees' perceptions surrounding SC are different. This study serves as a starting point for analysis, focusing on interpretation.

In summary, studies have shown that seemingly small or insignificant words or phrases in English can carry robust social meaning. We understand that much is communicated beyond what is explicitly stated, and (a la Grice 1975) participants in a conversation have expectations for what that conversation will entail and how it will unfold. Various features of speech and conversation can contribute meaning directly related to social factors (Campbell-Kibler 2009, 2010). These types of studies have been done on determiners and demonstratives (Acton and Potts 2014, Acton 2014, Acton 2019), fillers such as non-lexical sounds or word repetitions (Stubbe & Holmes 1995, Mata 2016, Martínez 2011), interjections (Norrick 2009), and other types of discourse markers such as *so* and gendered speech (Bolden 2009, Mata 2016, Cheshire 2005). These studies show the important social meanings that can be carried by seemingly small or insignificant parts of speech, providing motivation for the present experiment to pinpoint the social meaning communicated by SC.

3.2 Methodology

3.2.1 Design

The design of this experiment is loosely based off that presented in

Campbell-Kibler (2009), which focuses on manipulating a single variable to see if that small change is enough to license a different interpretation of stimuli among listeners. Campbell-Kibler's use of an experimental task accompanied by an interview/survey was able to capture and confirm participant judgements about the speakers they heard, which is why I chose these same types of tasks for this study. This study was situated in an interface designed to resemble a social media platform, which served to present the tokens of SC in an environment where participants had likely encountered it before, and also to make the participants as comfortable as possible, as social media is likely something they are familiar and at ease with more than a standard experiment format.

The specific purpose of this experiment was to elicit reactions to naturalistic usages of SC from native speakers of English. This was accomplished using two tasks. The first was a multiple choice reactions task in which participants read short, 1-2 sentence discourses containing either SC or other standard determiners (*the* or *that*), and selected from six emoji the one that best fit their reaction to what they read. The second task consisted of several open-ended questions about SC, designed to capture speakers' intuitions about how SC is used and what they believe speakers are trying to communicate when they use it. In this task, participants were provided with an example sentence containing SC, and asked if they were familiar with and had encountered *said* used in this way on social media. They were also asked on a Likert scale (ranging from never to frequently) how often they choose to use this construction themselves. Finally, this task included questions about demographic information such as age and self-reported estimated social media usage.

3.2.2 Participants

204 participants were recruited using prolific.ac, which has many available selection criteria for narrowing down an eligible participant pool. This study was made available to participants who were over 18 years of age and were born in and currently living in the United States. Table 3.1 below shows the distribution of participant ages for this study:

Age range	Number of participants
18-25	52
26-35	91
36-45	33
46+	28

Table 3.1: Participant ages, English study

Participants also had to report having spent the majority of their life and time in the U.S. While participants were not required to be monolingual English speakers, they were required to report English as their primary language and the language in which they conduct the majority of their day to day life. Out of Prolific's 40,000+ participant pool, 13,528 were eligible for the study. Participants took an average of 5 minutes to complete the study, and were paid $1.05 (or $12.60/hour). All 204 responses were collected on the same day, in July of 2018.

3.2.3 Task 1

The first task was designed to closely resemble Facebook, a platform which

many people use and are familiar with. The motivation for situating this study within a social media-type interface is threefold: First, since participants are likely familiar and comfortable with social media, this experiment design provided a natural environment in which for them to interact with SC in a way that closely resembles "real life"[12]. Second, since social media is a common medium for encountering tokens of SC, I was able to base my experimental stimuli off actual tokens of SC taken from social media, making the reaction data I collected as naturalistic as possible. Finally, this task served as a "fun" priming task for the second portion of the study by exposing participants to tokens of SC so that they were better prepared to think about and discuss their intuitions about it in an open-ended manner during task 2.

Sentences for the first task were adapted from actual tokens of SC I have collected, both from curated corpora (namely the data presented in chapter 2 from the Corpus of Contemporary American English) and "in the wild" from television, blogs, and social media platforms such as Facebook and Instagram. As explained in chapter 2, COCA was searched for instances of *said* tagged as an adjective by COCA's algorithm, followed by a noun ("said_j NOUN") or an adjective ("said_j ADJ"). These searches produced total 1013 and 91 tokens respectively, at the time of collection.[13] Tokens were read individually to ensure the proper usage of *said*, which was determined based on syntactic location (usually directly after a verb, but occasionally sentence-initial), and the presence of a different verb in the sentence, to ensure that *said* was not interpreted as

[12] Participants were asked to report their estimated social media usage. 68% of participants estimated spending at least 3 hours per week on social media, while only 6% estimated spending fewer than one hour per week.
[13] COCA frequently adds new data, so these numbers may be slightly larger now.

such. Additional tokens were gathered as encountered over the course of five years. Tokens were chosen for the experiment based on how "social media-like" they were, as determined by short length of sentence, and informal use and subject matter. Only tokens that were not overtly humorous were chosen; no tokens containing intentional jokes or recounting funny stories/occurrences were used.[14] Each sentence in the experiment had a duplicate counterpart featuring a standard determiner instead of SC, as in (1):

> (1a) We're adults, we bought a house, we also may or may not be currently playing hide and seek in **said house**.
> (1b) We're adults, we bought a house, we also may or may not be currently playing hide and seek in **that house**.

There were two experimental blocks each containing ten sentences, half of which were SC-containing sentences and the other half contained standard determiners[15]. The sentences in the experiment blocks were identical, varying only by use of *said* or a standard determiner such as *that* or *the*. The sentences were presented visually in a way that closely resembled Facebook; sentences were accompanied by a "profile picture" and a (fabricated) name of the assumed speaker/poster of the status. Social media profile photos were obtained either via stock photography or from friends and colleagues with permission from the individuals present in the photos, as approved by the IRB. These photos were cropped and sized down to a standard Facebook thumbnail size, approximately one square inch. Photos were chosen that contained individuals

[14] While it is possible that the "social media-likeness" by itself could be enough to license humor, the purpose was not to look at the humor of any individual token, but rather to assess the difference in humorous interpretation as caused by the change in determiner. In other words, the comparison of reactions between (2a) and (2b), which vary only by determiner is more important that the humorous interpretation of either (2a) or (2b).
[15] See Appendix B for a list of all stimuli questions.

that matched the potential speaker of the sentences based off of the information presented in the sentences. For example, two sentences in the study mentioned a baby; one was accompanied by a profile picture containing a family, and the other by a photo of a dog standing next to a newborn.

A popular trend for social media websites is to include a reactions/rankings poll at the end of each article. In task 1, reaction-type emoji were presented below each sentence based off the standard reactions provided on Facebook. Within the Facebook user base, each of the reaction choices has a perceived and accepted meaning by users of social media, and are used in Facebook algorithms to customize users' newsfeeds to include more material similar to that which they react most strongly (Moreau 2018, Constine 2016). These conventionally agreed upon meanings are shown in Table 3.2 below.

Emoji	Title	Use
👍	"Like" or "Thumbs up"	The original, sole reaction available on Facebook until the introduction of the rest of the reaction options in early 2016. This reaction denotes general acknowledgement and support.
❤	"Love"	A stronger reaction than the generic "thumbs up" and is used for expressing more enthusiasm or support than could be conveyed by a thumbs up.
😂	"Heehee"	Reserved for posts that readers interpret as humorous. It is the only reaction that clearly expresses laughter and amusement.
😮	"Wow"	Used for surprising posts and to express shock. This seems to be used when a "like" doesn't seem quite right but users still want to react.
😡	"Angry"	Often used to show disapproval over wrongful actions. This is often seen in response to controversial, political or news oriented posts.
😢	"Sad"	Used to show empathy over a sad post, or similarly to the angry emoji in reaction to upsetting or controversial stories.

Table 3.2: Meanings of Facebook reaction emoji. Descriptions based in explanations provided in Moreau 2018 and Constine 2016

Participants in the present study were instructed to read each sentence and interact with it as if they were on a social media site. Since each participant was presented with only ten sentences, this task took approximately 3 minutes to complete. All participants finished this task and response data was analyzed for relationships between the determiner the sentence contained and the emoji the participant used. I hypothesized that SC-containing sentences would be more likely to use a 😂 emoji when compared with the duplicate sentence using a standard determiner, which I predict would be more likely to use a standard reaction response, 👍. This hypothesis is rooted in the idea that the unexpectedness of once-formal *said* in an informal context could be interpreted as humorous in the discourse, based off Acton's Violations of Expectations principle. A block 1 sample question containing the standard determiner-containing NP *the newborn* is included below in figure 3.2, and an additional example is included in Appendix C.

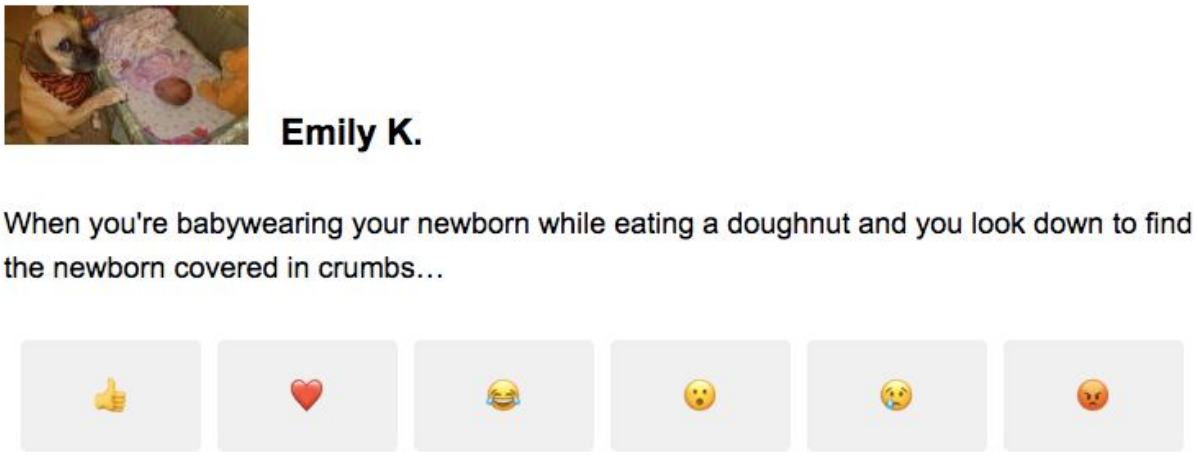

Figure 3.2: English task 1 sample question

Participants completing block 2 encountered this same question, but containing *said newborn* instead of *the newborn*.

3.2.4 Task 2

After completing the first task, participants were automatically presented with the second task. As stated above, this was a survey task that was comprised of both demographic information such as age and estimated social media usage per week, and open ended questions specifically about SC and their perceived usage of the construction.[16] Participants were asked if they are familiar with SC, if they regularly encounter the construction in daily life, and if they ever choose to use it themselves. These questions were presented in various formats, including fill-in-the-blank, short answer, and Likert scale questions. After having interacted with SC-containing sentences in task 1, the construction was fresh in the minds of participants, giving them an opportunity to share insights and intuitions about the construction. This task took an average of 2.5 minutes to compete, with an overall average time of 5.5 minutes (330 seconds) for the entire study.

3.3 Results

3.3.1 Task 1

For task 1, the breakdown of total emoji responses for SC-containing sentences and standard determiner-containing sentences are as follows, in Table 3.3:

[16] The questions asked in task 2 are available in Appendix F

	SC-containing Sentences	Standard D-containing Sentences
👍	34%	35%
❤	13%	14%
😂	30%	25%
😮	13%	15%
😡	5%	5%
😢	4%	5%

Table 3.3: Results of reaction emoji with determiner type, English study

From this raw data, almost all emoji categories are the same (within 1-2 percentage points) between determiner groups. The laughter/haha emoji is the only emoji that differs by a larger range, showing that participants are more likely to react to sentences containing SC with a humorous response than sentences that contain a standard determiner.

To test for statistical significance, this data collected from task 1 was analyzed using a general linear model, specifically focusing on the relationship between determiner type and emoji usage. In this model, the goal was to see if a 😂 reaction could be predicted by the type of determiner used in the sentence, by looking for a correlation between an SC containing sentence and the use of a 😂 emoji as opposed to this same emoji with a standard determiner containing sentence. This model was run controlling for participant, and with and without controlling for the specific question participants were answering. The effect was significant in both models, but slightly more significant when including the question as a random intercept. In this analysis, values such that $p<.05$ are significant. As shown in table 3.4, there was a significant effect

between determiner and reaction emoji used.

	Estimate	Std. Error	z value	pr(>|z|)
intercept	-1.0450	.2904	-3.598	.00032
Standard determiner	-.3092	.1088	-2.841	.00450

Table 3.4: Fixed effects for English SC social meaning experiment

Group	Variance	Std. Dev.
Participant (intercept)	.3804	.6168
Question (intercept)	.7619	.8728

Table 3.5: Variance and standard deviation for random intercepts in English SC experiment

These results show that sentences containing SC were significantly more likely to be rated with a 😂 reaction than sentences containing a standard determiner, with a value of [p<.001].

3.3.2 Task 2

In task 2, participants were presented with a series of open-ended questions for the purpose of eliciting feelings about SC that could potentially not be captured as clearly in task 1, where there was potential concern that the presence or absence of SC in a sentence would not be enough to elicit a change of reaction. Additionally, these questions were open-ended to prevent an effect from feeding descriptions to participants that may not match their own judgements, such as in a more standard attitudes test methodology.

The theme of this task revolved around participant familiarity with SC and

perceived social meaning. Of the 204 participants, 134 (67%) reported that they were familiar with and regularly encounter this construction in their normal social media usage. 9% of the remaining participants were unsure, and 24% did not believe they had encountered it before. Participants were also asked to select frequency options from a five-point Likert scale, denoting how often they themselves choose to use SC in their own social media postings. Only 6% of participants reported using the construction frequently or regularly, with the remaining 94% selecting sometimes (29%), rarely (37%) or never (28%). This shows that while the majority of participants regularly encounter SC on social media, few of them (claim to) choose to use it regularly themselves.

In one question, for example, participants were asked to complete the following sentence: “I think people use *said* when they want to seem _____.” They were provided with an empty text box and were instructed to write as little or as much as they wanted. Results were grouped into categories that matched any patterns that emerged from the data, as explained in the methodology. Some participants chose not to answer this question, while others provided multiple adjectives and/or answers. Each adjective was counted individually, so the number of responses is greater than the number of participants. Patterns were very evident, with the majority of participants responding with *funny* or similar words such as *silly* or *witty*, or *intelligent* and related words like *smart*, *educated*, or *well read*. Most responses fell neatly within 7 categories:

Category	Examples of responses included in category
Funny	silly, witty, humorous, quirky
Intelligent	smart, educated, well read, knowledgeable
Formal	fancy, sophisticated, mature, pretentious

Irritated	annoyed, bemused, irked
Sarcastic	snarky, ironic
Hip	cool, extra
Clear	emphatic, specific, precise
“Other”	coy, different, stupid, focused

Table 3.6: Category distributions of adjectives used to describe SC

The “other” category contained words that did not seem to fit into any of the other categories, and several responses were excluded for either being left blank, or for having answers such as “I don’t know” or similar.

Results of each category were totaled, and are shown below in Figure 3.3.

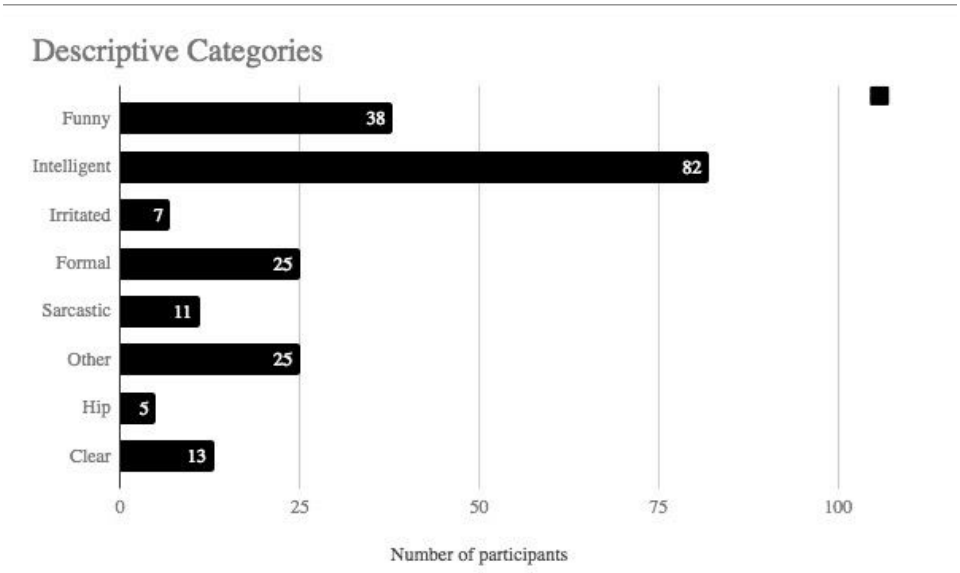

Figure 3.3: Totals of adjective categories

58% of total responses stated that people use SC when they want to seem either *funny* or *intelligent*. The category with the largest number of responses were words related to

intelligent such as *smart* and *educated.* The second most used adjective was *funny* or related words such as *witty* and *silly,* comprising 18% of responses. Around ten participants stated both of these adjectives, and interestingly, many of the participants who reported that people may use SC to sound smart or intelligent also added a clause that these speakers "do not take themselves too seriously," showing that there is an underlying humorous component even to a *smart*-sounding usage of SC. The category marked *other* included uncategorizable terms such as *dramatic*, *weird*, *whimsical*, *unique*, *emphatic*, *interesting*, *stupid*, *normal*, and *basic.* While these words did not fit into any of the aforementioned categories, they do support the idea that participants believe that speakers have a purpose in mind for choosing to use SC over a more standard determiner form[17]. Participants had the option to leave this answer blank or answer with "I don't know", "nothing", etc, but only three participants chose to do this. I believe more participants would have chosen not to answer this question if they believed SC's function in discourse was similar to that of a standard determiner, or lacking an external meaning. By choosing an adjective at all, participants are pointing to the new, developing meaning of *said*.

One problem I encountered with the open-ended question format was that certain descriptive words participants used for question 6 (see Appendix F) could potentially fall into more than one category. For example, several participants used the adjective *fancy* in their answer, which I ultimately grouped in the *formal* category. I do, however, acknowledge the argument that one could easily consider academic speech fancy, and I

[17] It should be noted that some participants chose not to answer, or answered with "I don't know" or "nothing". These responses were not included in the analysis.

grouped any references to education and academics in the *intelligent* category. Another problematic case was *witty*, which was an adjective used by several participants, and could be argued to mean both funny and intelligent, but was ultimately counted in the *funny* category. I did not group any answer to this survey question into multiple categories unless a participant had a qualifying phrase after their adjective. For example, one participant stated that they believe people use *said* when they want to seem "educated, like they are trying to sound a little more fancy by spicing up their sentence." This answer was tallied in both the *intelligent* and *formal* categories.

In addition to this "fill in the blank" style question, participants were given an essay-style question inviting them to share intuitions about why people would choose to use *said* in this way, and if they believe it has a specific purpose. Answers varied widely for this question, from "I don't know" to "to refer to something they already mentioned". Some of the more interesting and potentially informative answers are included below.

- "It's funnier than *the baby* or *the donut*. It reads as more emphatic and adds an air of bemused frustration." -Participant 10
- "They think they sound wiser and more intelligent." -Participant 15
- "It is any easy way to add formality to a simple statement in a way that conveys humor." -Participant 18
- "I think they are trying to mimic the sound of a cheeky, personal diary" -Participant 3
- "It is a way to casually emphasize a target noun that was already brought up. Usually for humorous effect." -Participant 49

- “It signals a joke usually, but like a witty or intellectual one” -Participant 95

These answers speak to the idea of social meaning, showing that SC is interpreted in a specific way, and conveys meaning that is not communicated through the use of a standard determiner.

3.4 Discussion

In summary, participants of this experiment reacted to social media style “posts” using the standard reactions schema provided on social media sites such as Facebook and Buzzfeed. They also answered open-ended questions about SC aimed to elicit answers providing information about their understanding and interpretation of the construction. The goal was to see if SC is interpreted as having a particular meaning or communicative function in discourse, and if individuals respond to reading or hearing it differently than they would a standard determiner.

Results showed that SC containing sentences are more likely to be reacted to with a response indicating a humorous interpretation by individuals, at a statistically significant level when compared with identical sentences containing a standard determiner. Furthermore, participants were most likely to express that they believe SC is used to denote a type of mock-formality in discourse; many participants acknowledged that it is formal, yet with a humorous or unserious twist. These types of answers support the idea that *said* carries a social meaning and may be intentionally employed by speakers to accomplish this conversational goal.

The main purpose for creating an experimental methodology centered around social media was to make this study as naturalistic as possible both in terms of how the stimuli were presented, and in how the participants interacted with it. The Said Construction can be readily observed in social media, and with Americans spending an estimated 40 or more minutes per day on Facebook, it is highly likely that participants encounter SC during their normal social media usage (Frier 2014). With this in mind, I created an experimental interface that allowed participants to complete a task that most likely looks like something they already spend a lot of time doing on their own.

The ten questions presented in each block of Task 1 were based off of actual tokens of SC collected from popular social media sites such as Facebook, Instagram, and Twitter, as well as web-based tokens from the COCA. 200 participants completed this task (100 for each block). The main goal of this task was to see if a change in determiner was enough to elicit a different reaction across blocks. Sentences in block 1 were identical to that of block 2 other than the determiner, which was either standard (*the*, or *that*) or *said*. I hypothesized that if SC does convey a humorous or mock-formal social meaning, sentences containing it should be reacted to differently than sentences containing a standard determiner. This was reflected in the data, which showed that SC-containing sentences were significantly more likely to receive a humorous reaction than standard determiner-containing sentences, which were more likely to receive a standard “like” or “love” reaction. While there was a significant effect between determiner type and emoji reaction, there was surprisingly not a significant effect between the age of the participant and the likelihood that they would use the 😂

reaction. This could likely be due to the distribution of the ages of participants; 75% of participants were under the age of 35. Further studies with more balanced age groups would be needed to account for a potential age effect.

For Task 2, I predicted that participants would use words like “humorous” to describe the meanings they believe speakers have in mind when using SC. It was surprising to find that words related to “intelligent” were the most common in terms of participants’ perceptions of why speakers choose to use SC, and also interesting that many participants who responded using “intelligent” qualified it by saying that it had an underlying humorous element. I also predicted that few participants would report using this construction frequently, and that the majority would report using SC “sometimes” or “rarely.” This was reflected in the data, which showed that only 6% of participants selected the two highest frequency categories for SC usage. Even so, 67% of participants claimed to be familiar with SC and reported that they regularly come across it in their normal social media usage, which could suggest that perhaps the few individuals who do choose to use SC do so regularly, and those who encounter it perhaps see it from the same few speakers repeatedly. Interestingly, there was no correlation between participant familiarity with SC and reported social media usage; participants who spent 10+ hours per week on social media were no more likely to report being familiar with SC than those who spent between 1 and 3 hours.

The most noteworthy observation from this study is that the majority of answers about the perceived meaning of SC (120, or close to 60%) stated that speakers’ use of SC conveys humor or intelligence. I believe that this provides evidence for the shift in

meaning that has happened over the past few decades, from *said* as a formal adjective meaning “aforesaid” or “aforementioned” to an informal determiner used to flag conversational significance of humor or mock-formality. This claim is operating under the assumption made by Tiersma (1999) that SC was once common in legal language but has become obsolete. If we ignore Tiersma’s claim and assume instead that SC is still commonly used in legal or formal discourse, we can use this data as evidence that SC in certain contexts--namely informal, web-based discourse--is used for a different purpose and to communicate a different meaning (humor and/or mock-formality) than SC in formal contexts. As shown through this project, this informal usage of *said* is popular in genres of discourse that people frequently use, such as social media, blogs, and other interactive web-based discourse, perhaps perpetuating this informal usage and social meaning. I believe this interpretation goes beyond just the choice of determiner, and is partially dependent on the expectations this construction carries. People are likely aware of SC’s history as a more formal construction, as evidenced by participants’ responses in task 2, and the associations with humor and intelligence are a result of both this history and the unexpectedness/unlikeliness of seeing *said* outside this traditional context.

While one could argue that perhaps this experiment only captured that SC is interpreted as having an intelligent or humorous social meaning on social media and that perhaps it is not necessarily generalizable to other genres of discourse, I do not believe this is a problem. Based on the corpus analysis presented in chapter 2, web-based discourse seems to be the genre in which SC is most commonly found

today, and I believe it is more likely that this social media-based usage is influencing how and if it is used in other genres, and not the other way around. It is safe to say that most people are more likely to spend time on their social media sites than reading or listening to legal discourse, and are using and interpreting language in a way that is conventional in their discourse communities, online and otherwise. This experiment allowed participants to interact with SC in the way they are most likely to encounter it, eliciting data that gives us an accurate picture of how it is used and interpreted in this moment in time. This picture may not have been accurate 20 years ago, and perhaps will not be so 20 years in the future, but I believe it is a solidly clear depiction of how it is being used and interpreted right now. Participants' intuitions that SC is "funny" or "intelligent" also seems to hint at the idea that people recognize the formal roots of this construction, and that the humor lies in the fact that this once-formal construction is being used in a context where it is unexpected. The unexpectedness of seeing SC in a casual context is perhaps why it is funny (or perceived as such).

This social media-centered methodology could easily be adapted to investigate other types of linguistic phenomena, especially those dealing with questions of how language is used and interpreted in this now common form of discourse. For example, this would be an effective methodology for looking at language within a social group via social media group pages, and comparing usage and social meaning with speakers outside of that social group. It is an unobtrusive way to observe language in real time and practice. Social media has become an invaluable tool for observing language patterns and seeing change happen; it is only natural that sociolinguistic

experimentation methods should adapt to make use of this resource. While this experiment was modeled after Facebook and used Facebook's default reactions bar, it would not be far-fetched to create a novel social media type experiment platform with reactions specifically tailored for answering specific questions. Since many people spend multiple hours per week on social media, these types of experiments are easy for participants to use and understand, and can provide compelling data for questions that can be answered without acoustic data.

CHAPTER 4

Comparative analysis of the Dicho construction in Spanish

4.1 Introduction

Similar to the Said Construction in English, Spanish has the Dicho Construction. [18] On the surface, SC and DC appear almost identical; they share Latin roots[19] (Norberg 1980, see section 2.2), appear in the same position syntactically directly preceding a noun and often right after a verb, and are both used to refer to something that has been mentioned or can be considered *given* in the discourse. A more in-depth analysis of *dicho*, however, shows differences in how the construction is used and interpreted between the two languages, such as the types of information with which they are most commonly used, and how they are interpreted by listeners or readers. This chapter aims to provide an overview of DC in Spanish both to draw a line of distinction between the two languages' uses of the construction syntactically, and to examine how DC is used and interpreted by speakers of Spanish. Similarly to chapters 2 and 3, this chapter investigates whether or not Spanish has maintained this construction as a more formal one, and seeks to find evidence that DC has undergone the same type of social change that SC has in English, and if it is also a carrier of social meaning.

[18] It should be noted that other languages have this type of construction too; French uses *le dit*, for instance. This analysis focuses only on Spanish and English, as explained and motivated in section 4.2 below.

[19] The constructions themselves share Latin roots, although the actual word *said* is of Germanic origin.

This chapter is comprised of two studies of Spanish *Dicho* that present methodologically similar, related analyses to the English SC studies presented in chapters two and three. I will show the results of a corpus-based study looking at patterns in usage and information status of the Dicho Construction, as well as a Spanish social meaning experiment closely resembling the one discussed in chapter 3. These studies together will provide evidence that while Spanish does have and widely use the Dicho Construction, it is used and interpreted differently by its speakers and addressees than SC is in English. Results of these studies, when compared to English, highlight these constructions' divergence in these languages from their originally similar usages in formal and legal language.

The first part of this chapter will provide an overview of DC, using corpus data from the Corpus del Español. Results of this corpus study suggest a few key differences between DC and SC, including higher frequency of use in DC, and a higher likelihood of DC to be used with information I categorized as inferred in chapter 2, or that which has not already been explicitly mentioned. From there, I will present the results of a social meaning experiment showing that DC is not widely interpreted to have a specific social meaning, unlike SC in English.

4.2 Corpus Analysis

The data for this analysis was collected using the Corpus del Español (CdE), which is housed under the BYU umbrella of corpora along with the COCA and other corpora used for the analysis presented in chapter 2 (Davies 2008). The CdE has several sub-corpora; the web/dialect corpus contains 2 billion words and is made up of

texts from millions of websites from 21 different Spanish speaking countries covering a timespan of about four years. The genre/historical corpus contains 100 million words of data from the 1200s-1900s. While this is not a fully balanced corpus, the data from the 1900s is balanced across four of the five genres present in the COCA: spoken, news, academic, and fiction. Finally, the News on the Web (NOW) corpus contains over 7 billion words of data from 2012-present, taken from internet news and magazine websites. For the sake of this study, only data from the web/dialect corpus was used, as there were several drawbacks to the genre/historical corpus: First, it is only balanced across genre for the 1900s, and does not have additional data from the 2000s. It also was only possible to take a random sample from the entire corpus, meaning a potential sample would not be balanced for genre. Finally, since many of the tokens in this corpus are historical and not web based, there is not a way to reliably track down antecedents for DC tokens if the antecedent is not available in the context provided by the corpus. Additionally, a drawback to potentially using the NOW corpus is that only contains news based sources, and in order to better frame the social media study, it was more natural to use web-based tokens for this analysis.

This analysis specifically focused on dialects of Mexican Spanish, so data collected from the web/dialect corpus was set to exclude tokens from other Spanish speaking countries. This was done primarily to mirror the English study as closely as possible; since the English study presented in chapter 2 focused only on varieties of US English, I wanted to limit the Spanish study to one country as well. Mexico was selected due to both its physical proximity to the US and also because it is the country with the

largest number of native Spanish speakers worldwide; approximately 25% of the total world population of L1 Spanish speakers are native to Mexico (Lyons 2017).

4.2.1 Methods

To obtain data for this analysis, identical search terms as presented in chapter 2 were used; the web-based portion of the CdE was searched for instances of “dicho_j* NOUN” and “dicha_j* NOUN”. While set to include lemmas, the results did not include tokens of the plural form *dichos*. Additionally, extra searches were needed to account for the adjective system of Spanish in which the adjective is generally placed after the noun but can occasionally be used before it. To account for these types of sentences, the searches “dicho_j* ADJ NOUN” and “dicha_j* ADJ NOUN”, accounting for instances of dicho/a followed by an additional adjective before the target noun. Finally, “dicho_j* NOUN ADJ” and “dicha_j* NOUN ADJ” were searched to find out how many NP tokens from the initial search contained a post-nominal adjective. “Dicho N” was also searched, as the majority of instances of *dicho* in the corpus were tagged as verbs instead of adjectives (which was not the case with *dicha*, as it is never used as a verb in Spanish), making the search results from the original adjectival search terms minimal.

4.2.2 Results

The initial searches collectively resulted in nearly 300,000 total tokens of DC, which included both grammatical genders (dicho/a) as well as instances with an intervening adjective. There were 132,163 instances of Dicha N, and 144,373 Dicho N. In searches involving an intervening adjective, it was slightly more common with NPs

using a masculine noun; there were 542 tokens of Dicho Adj Noun and 490 of Dicha Adj Noun. Table 4.1 provides a breakdown of the types of each response:

Search	Total
Dicho N	144,373
Dicha N	132,163
Dicho N Adj	6930
Dicha N Adj	6888
Dicho Adj N	542
Dicha Adj N	490

Table 4.1: Corpus results for instances of dicho tagged as an adjective, Spanish

A random sample of 200 tokens each (400 total) from the "dicho N" and "dicha N" results was taken, and a native Spanish speaking research assistant was trained to categorize and code responses for information status in accordance with the English study[20]. As the corpus consists entirely of web-based data, links are provided in the search results to the site from which the data was taken. The research assistant followed each of these links to read the expanded context of the token in order to correctly categorize each item. Since the *dicho* tokens were not tagged as an adjective, tokens using *dicho* as a verb were also excluded. Furthermore, since websites often undergo revision and change, many of the tokens in the sample were no longer available due to dead links or modified pages. With these exclusions, there were initially

[20] Initially, 400 tokens of "dicha N" were taken for this sample. Due to lack of availability from the research assistant, however, only half of the sample was finished and able to be used in the study. The "dicho N" tokens were coded by myself, with problematic cases clarified by a native Spanish speaking family member.

334 tokens of DC. The Internet Archive was used to check for records of the dead links, and the first 17 successful results were coded and saved, bringing the total number of tokens for this study up to 351, which was the number used in the English study. Identically to the study on SC, tokens were coded for presence or absence of an additional determiner and presence or absence of an adjective contained within DC, in addition to information status type categories. As explained in chapter 2, the type categories are broken down as follows: type 1 is for instances where the noun in the DC is the same word as its antecedent. In other words, it must be explicitly mentioned. For example, *la organización → dicha organización* (*the organization → said organization*). Type 2 is cases where the DC noun and its antecedent form a subtype/type relationship, or where the antecedent includes a property of the noun, but this property is not carried over into the DC, as in *una auditoría externa → dicha auditoría* (translated as *an external audit → said audit*, instead of *said external audit)*. Type 3 is for antecedents and nouns that are roughly synonyms of one another, such as *una regulación → dicha normatividad* (translated as *a regulation → said rule*). Finally, type 4 was cases where the noun can be inferred from the antecedent, such as *se contrapongan a la Carta Magna → dicha materia* (translated as *oppose the Magna Carta → said subject*). The breakdown of tokens in each category are as follows, in table 4.2:

Category 1	n=116 (33%)
Category 2	n=133 (38%)
Category 3	n=16 (5%)
Category 4	n=86 (25%)

Table 4.2: I.S. Results, Spanish

In the English analysis, 89% of tokens were in the first two categories, pointing towards SC's tendency to be used with information that has already been linguistically mentioned in the discourse. This tendency holds in Spanish, although it is not as prevalent, as only 71% of corpus tokens fell into these categories. Surprisingly, a large portion of Spanish data used DC in an inferential construction, which was the least common category in the English study; in the English study, only 13 tokens (4%) were type 4, but 86 (25%) of the Spanish tokens used *dicho/a* inferentially.

Category	English	Spanish
1	178	116
2	135	133
3	25	16
4	13	86

Table 4.3: Comparison of English and Spanish across categories

As stated above, this corpus contains data only from online sources from the past several years. Because of this, tokens were not coded for year and genre as they were in the English study, meaning diachronic and genre analyses were not possible. It should be noted, however, that the vast majority of tokens in this Spanish sample were from websites for the following categories: political candidates/campaigns, websites for cities or government agencies, and news sites. There were also a number of blogs from

sites like blogspot.com (as there also were for the English study) but there was a distinct stylistic difference in these blogs between the two language groups. In the English study, they were mainly “lifestyle” blogs; these types of blogs have become common over the past 10 years and are mainly a personal blog highlighting the author’s family, hobbies, cooking habits, clothing, etc. The blogs in the Spanish study were more topical or group-oriented, focusing on things like agriculture, law, or a social group. One specific example of this is a blog for the 1990 *Instituto National de Panamá*, which was used by members of this group to share photos and memories of their time together. The purpose of drawing this distinction between the types of sources from which the tokens came in Spanish and English is to begin to highlight the idea that these constructions are used differently in these languages and are found in different types of discourses.

4.3 Social Meaning

The purpose of the second Spanish study, on social meaning, is to investigate the perceived motives and interpretations of DC by native Spanish speakers. It is rooted in the assumption that speakers of Spanish perceive and use this construction differently from English speakers, as evidenced by its higher likelihood of being used with inferred information and also its tendency to appear in discourses that are more formal than where SC is typically found in English. If this assumption is correct, and if the random sample collected from the Corpus del Español is representative (for the purpose of this analysis, we are operating under this assumption), one would expect

Spanish speakers to interpret this construction as being more formal in nature, and perhaps lacking the social meaning that was evident from the English social meaning study discussed in chapter 3.

The experiment presented in chapter 3 used a social media type interface to elicit reactions from participants on sentences containing SC vs. a standard determiner such as *the* or *that*. This chapter presents the results of an analogous study conducted on DC with native Spanish speakers. While to my knowledge there are no existing studies on *dicho* (syntactically or sociolinguistically), there are several studies that have looked at the social meanings of various grammatical features in Spanish. Mata (2016), for instance, investigated the use of fillers such as *so*, *okay*, *eh*, and *pues* in various demographics of Spanish speakers in the San Diego/Tijuana border region. Mata looked at residents on both sides of the border including 1st and 2nd generation immigrants to the San Diego area and found that the use of *so* as a filler is perceived as *americano* by monolingual Spanish speaking listeners who have not lived north of the border (Mata 2016, 161). Another study (Chappell 2016) looked at the voicing of intervocalic /s/ in Costa Rican Spanish, which results in phenomena such as *pasa* (raisin) being pronounced [paza]. Participants in this study were more likely to associate tokens containing the intervocalic [z] with lower social status, but were also more likely to rate these tokens as higher than [s] for niceness and, according to male participants, masculinity.

4.3.1 Study design: Task 1

To the extent possible, the stimuli for this experiment were identical to those of the English study. Items were translated to Spanish by a native Spanish speaker who was born and raised in Mexico and moved to San Diego, California as a teen. She has native-like fluency in English, and has completed rigorous training to become a court translator. Given her training, she is skilled at translating nuances and maintaining the intended meaning from English. Where necessary, items were changed to be more culturally relevant for a Spanish speaking audience. For example, in the English study, one experiment question was about sweet potatoes:

> (1)“So... in theory I like sweet potatoes, but said theory includes loads of brown sugar, butter, and marshmallows.”

The translator was concerned that this sentence would not translate in a way that would be relevant to Spanish speaking participants because sweet potatoes are not part of the standard Mexican diet. Instead, she developed a similar sentence using *plantano macho*, which is a dish consisting of plantains that are fried and eaten with sugar and cream:

> (2) “En teoría me encanta el plátano macho, pero dicha teoría incluye que se fían, mucha azúcar y crema.”
> [In theory I like plántano macho, but said theory includes lots of sugar and cream.]

As discussed in chapter 3, the study is designed to resemble a social media platform, with profile pictures and names accompanying each sentence. For the Spanish experiment, names of some of the fictional social media participants (in other words, the “speaker” of the sentence) were changed to more Spanish-sounding names, or ones that are commonly found in both English and Spanish, such as *Jaime* and *Adriana*.

Stimuli examples for both the English and Spanish versions of this experiment are available in Appendices D and E. Also as discussed in chapter 3, an emoji-based reactions bar was provided to participants as a tool for them to use to interact with the experiment stimuli. Again, each of the reaction choices provided by Facebook has a perceived and accepted meaning by users of social media, and are used in Facebook algorithms to customize users' newsfeeds to include more material similar to that which they react most strongly (Moreau 2018, Constine 2016). Table 4.4 repeats the description of the common use and meaning of each emoji in this reactions schema as presented identically in table 3.2 above.

Reaction Emoji	Title	Use
👍	"Like" or "Thumbs up"	The original, sole reaction available on Facebook until the introduction of the rest of the reaction options in early 2016. This reaction denotes general acknowledgement and support.
❤	"Love"	A stronger reaction than the generic "thumbs up" and is used for expressing more enthusiasm or support than could be conveyed by a thumbs up.
😂	"Heehee"	Reserved for posts that readers interpret as humorous. It is the only reaction that clearly expresses laughter and amusement.
😮	"Wow"	Used for surprising posts and to express shock. This seems to be used when a "like" doesn't seem quite right but users still want to react.
😡	"Angry"	Often used to show disapproval over wrongful actions. This is often seen in response to controversial, political or news oriented posts.
😢	"Sad"	Used to show empathy over a sad post, or similarly to the angry emoji in reaction to upsetting or controversial stories.

Table 4.4: Meanings of Facebook reaction emoji. Descriptions based in explanations provided in Moreau 2018 and Constine 2016

4.3.2 Participants

As in the English study, 200 participants were recruited from prolific.ac. Restrictions were set such that the study was limited to native speakers of Spanish who consider Spanish their first language and the language in which they conduct the majority of their daily interactions. Participants were born in and currently living in either Mexico or the US. Initially these restrictions were set to make the study available only to those born in and having lived their lives in Mexico, but this narrowed down the participant pool to fewer than 200 eligible participants. Adding eligibility to Spanish speakers living in the United States increased the participant pool to a larger number, ensuring 200 study participants. As stated, however, all of these participants reported Spanish as their first and primary language of use, hopefully minimizing any contact effects that would be present due to opening the study to Spanish speaking residents of the United States. In this study, the age ranges of participants were as follows:

Age range	Number of participants
18-25	102
26-35	74
36-45	20
46+	4

Table 4.5: Spanish study participant ages

All participants were asked to estimate their social media usage (across all social media platforms) using a likert scale[21]. 60% of participants reported spending at least 6 hours

[21] See Appendix G

per week on social media. Table 4.6 below illustrates their self-reported social media usage.

Estimated hours	Number of participants
0-1	3 (2%)
1-3	24 (12%)
3-6	53 (27%)
6-9	58 (29%)
10+	62 (31%)

Table 4.6: Spanish study participant self-reported social media usage

4.3.3 Task 1 results

The breakdown of total emoji responses for SC-containing sentences and standard determiner-containing sentences are as follows:

	DC-containing Sentences	Standard D-containing Sentences
👍	32%	34%
❤	12%	12%
😂	27%	28%
😮	17%	14%
😡	4%	5%
😢	7%	7%

Table 4.7: Results of reaction emoji with determiner type, Spanish study

These categories were largely the same between determiner types, with the largest

difference in total responses being the 😮 (*wow*) emoji, which had slightly more (although not significant) responses for DC-containing sentences than standard determiner containing ones. And surprisingly, DC-containing sentences had *fewer* overall 😂 (*haha*) responses than ones that contained a standard determiner. In other words, more participants responded humorously to standard determiner containing sentences than DC-containing sentences (although again not with statistical significance).

For the sake of continuity between the English study and the Spanish one, I tested for a relationship between the use of a *haha* emoji with a DC-containing sentence. A general linear model fit by Laplace Approximation was used, with the participant and the question set as random effects, which in the English experiment made the already significant effect stronger. In this Spanish study, however, there is no significant relationship between the participants' use of the *haha* emoji and the determiner type contained within the sentence. In table 4.8 below, the *haha* emoji is the intercept, and we see there is not a significant effect between its use and the type of determiner in the question sentence. Table 4.9 further shows no significant effect between any individual participant or question and the likelihood of a *haha* reaction.

	Estimate	Std. Error	z value	pr(>|z|)
intercept	-1.15122	.30024	-3.834	.000126
Determiner type	-.02914	.10903	-0.267	.789281

Table 4.8: Fixed effects for Spanish DC social meaning experiment

Group	Variance	Std. Dev.
Participant (intercept)	.08521	.2919
Question (intercept)	.83099	.9116

Table 4.9: Variance and standard deviation for random intercepts in Spanish DC experiment

Furthermore, as in the English study, there was no significant relationship between participant age and likelihood to react to a DC-containing sentence with any particular emoji option.

4.3.4 Study design: Task 2

A survey task[22]analogous to the English one was translated using the same native Spanish speaking research assistant who translated stimuli for task 1. All 200 participants completed this task, and were presented with the same questions and example sentences as in the English study. As described in chapter 3, this task consisted of questions about the Dicho N construction, including participants' familiarity with the construction, whether or not they have used it themselves, and what they believe it means or functions to accomplish in the types of sentences in which it is used. These questions ranged in response types with Likert scale, fill in the blank, and open ended essay-type questions.

4.3.5 Task 2 results

The main purpose of this task is to delve into the question of whether there is any sort of social meaning carried by this construction. In chapter 3, I showed that English speakers interpret SC to have a meaning of intelligence with a bit of humor or wit.

[22] Survey tasks for English and Spanish are available in Appendices F (English) and G (Spanish)

Results from task 1 show that it is not the case that DC is interpreted as being humorous by Spanish speakers. The survey questions in task 2 aimed to support that data by eliciting participants' attitudes and opinions about the construction.

In terms of study demographics, Spanish participants were more likely than English participants to self-report that they use DC sometimes, regularly, or frequently; 95 Spanish speaking as opposed to 70 English speaking participants reported using this construction. This information is provided in table 4.10.

Frequency	Spanish	English
Never	17	56
Rarely	88	74
Sometimes	76	58
Regularly	15	10
Frequently	4	2

Table 4.10: Participant self-reported use of DC and SC, study comparison

First, participants were provided with an example sentence containing the *dicho* construction, and were asked if they have encountered *dicho* used in this way, as opposed to the standard verb form. Only 70 participants said yes (compared to 134 in the English study), with the remaining 130 saying no (121) or that they were not sure (9). They were also asked if and how often they choose to use this construction, on a likert scale ranging from never to frequently, as shown above in table 21.

Question five in the survey asked participants why they believe people use *dicho* in this way, and if they believe it serves a specific purpose. As you will recall,

participants in the English study had strong feelings about this, and many provided answers that mentioned humor or wittiness in some form. For this question in the Spanish study, not a single Spanish participant provided any reference to humor in their answers; most participants responded that it is only for emphasis, or that it is formal (or a combination of both). Some participants claimed that this use of *dicho* is very common, while others stated that it is rare. Still more made a point that it is common to see this construction in more formal language, but it would be rare to use it or see it in social media. Others believe it has no purpose at all. Of the 200 responses, 28 said that they believe it serves the purpose of referring to something you have already mentioned while allowing the use a different word to describe it, for the purpose of avoiding redundancy in the discourse. The information in table 4.11 shows a representative sample of the most common types of answers:

Spanish	Translation
Una manera más formal de decir las cosas, nada más.	A more formal way of saying things, nothing more.
Para no repetir palabras y tratar de sonar más[23] interesante	To not repeat words and try to sound more interesting
Lo usan para evitar malentendidos[24] y hacer énfasis a lo que se refieren	They use it to avoid misunderstandings and to emphasize what they refer
La gente utiliza algunas palabras de una manera extraña, ya que no saben el uso correcto de estas palabras, sobre todo en redes sociales.	People use some words in a strange way, since they do not know the correct use of these words, especially in social networks.
Para referirse a algo mencionado anteriormente y no tener que volver a repetirlo.	To refer to something mentioned above and not have to repeat it again.

[23] Accent mark added; participants' original answer said *mas*
[24] Spelling corrected from participants' answer, which said *mal entendidos*

Table 4.11: Perceived purpose for DC, Spanish study

Question six, a fill-in-the-blank style question, asked participants to complete the following sentence: “Yo pienso que la gente usa “dicho” cuando quieren parecer _____.”[25] As in the English study, responses were grouped together by like words, and categories were formed based from these groups. For instance, words like *intelligent*, *educated*, and *smart* were grouped together into a single category, as were words like *strong*, *forceful*, and *intense*. For this question, answers outside my scope of Spanish ability were translated using Google Translate, and these translations were approved by a Spanish speaking research assistant. Participants were instructed to say as much or as little as they wanted, so responses with multiple adjectives were separated into different categories as necessary. Responses related to “I don’t know” or “nothing” were excluded from this analysis. The chart in figure 4.11 shows the frequency of the most common responses among Spanish speakers.

[25] English: “I think people use “said” when they want to seem _____.”

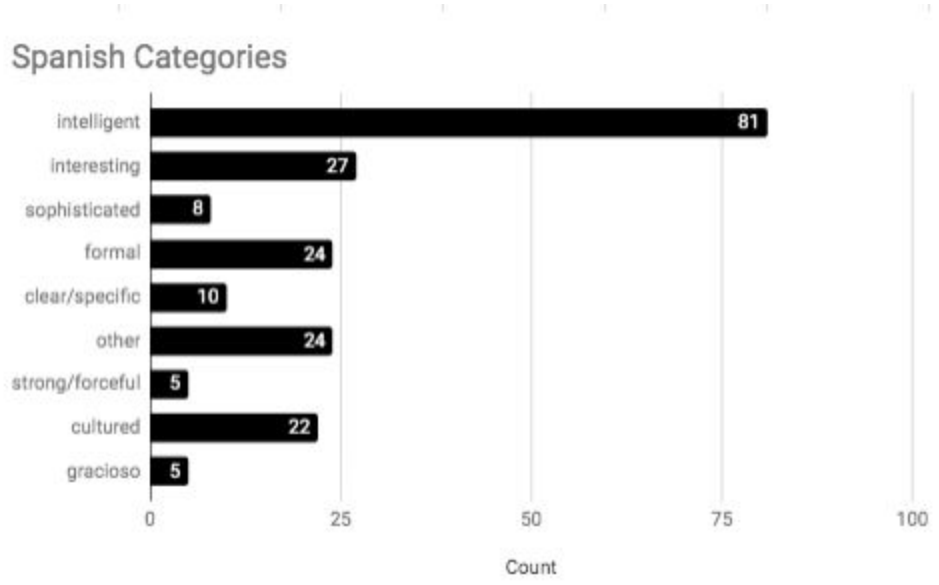

Figure 4.1: Response categories, Task 2, Spanish study

Responses for this question varied somewhat from those in the English study; while popular responses for both languages included categories for *intelligent* and *formal*, a large number of Spanish speaking participants used the word *interesting* to describe a speaker's motivation for using this construction. Only one or two English participants chose this word. Another popular response among Spanish speakers was *cultured*, which was not used by English speaking participants at all. Most of the Spanish responses centered around the ideas of formality and intelligence, whereas responses in the English study were thematically more humorous, with words like *funny*, *witty* and *sarcastic*.

For this question, there was also a tendency for participants to qualify their answers with *muy* or *más*. Table 4.12 shows several examples of this:

Spanish	Translation

Más inteligente de lo que es	More intelligent than they are
Mas imponentes	More impressive
Más cultos de lo que son	More cultured than they are
Más interesantes de lo que son	More interesting than they are
Mas específica de lo usual	More specific than they usually are

Table 4.12: DC Task 2, examples of answers using *más*

Of the 200 participants, 25 of them (12.5%) used *muy* or *más* in their answers. Based on answers like the ones presented in the table above, this seems to show that there are feelings among speakers that the use of this construction is "over the top" or excessive. Although answers like "mas inteligente" were categorized with other answers for "intelligent" (or "más cultos" with "cultured", etc), the use of *más* does seem to add another layer of meaning, and from these types of answers, we see that DC is not always interpreted positively. There is a sense that some addressees may interpret this construction as obnoxious, pretentious, or a sort of linguistic "eye-roll". Interestingly, none of the participants who stated that they use DC regularly or frequently provided these types of *muy*/*más* answers for its meaning. Instead, frequent users of DC used words like *claro* (clear), *ortográficamente correctos* (orthographically correct), and *conciso* (concise) when speaking for its meaning. All of the participants who provided answers using *más* or *muy* said that they use DC either never, rarely, or sometimes. This shows that there is a sort of dissonance between how the construction is used and interpreted among different groups of speakers. It was not clear, however, what may be a dividing factor between these groups of speakers; similarly to the lack of correlation

between age and emoji reaction type, there was no relationship between participant ages and likelihood to use *más* in an answer. This supports the notion that this is not necessarily a construction that is interpreted differently between older and younger speakers.

One interesting and potentially problematic case in the Spanish study was the word *gracioso*, which was used by five participants. This word can have a range of meanings in Spanish, from *funny* or *humorous* to *graceful*, and it seems that participants may be expressing this full range of meanings given their differing explanations for what they believe *dicho* means and why it is used. Consider in table 4.13 the following three responses to question five[26], which were all answers provided by participants who used *gracioso* as their answer to question 6, the fill-in-the-blank question:

Question 5, Spanish answer	Rough Translation
Tomando el ejemplo probablemente quieran verse con un vocabulario un tanto más refinado según ellos.	The speaker wants to be seen as having a refined vocabulary.
Para expresarse de forma más sucinta	To express something more succinctly.
Es una expresión o palabra más propia, en cierto sentido. También sirve para enfocar la atención en los objetos de otra forma, he visto la figura utilizada en chistes muchas veces.	It is a more proper expression or word, in a sense. It also serves to focus attention on objects in another way, I have seen the figure used in jokes many times.

Table 4.13: Perceived purpose of DC by users of *gracioso*, Spanish study

These answers are broad in scope; the first answer could be interpreted as meaning "refined" or "sophisticated", but the fact that this participant also used *gracioso* could

[26] See appendices F (English) and G (Spanish)

provide evidence that this answer may have an underlying meaning of “wants to be seen this way, but is actually seen as being humorous, or can’t be taken seriously.” When considering the second answer, this participant’s use of *gracioso* could be interpreted to mean something closer to the *graceful* or *sophisticated* meaning. And in light of the third answer, *gracioso* could be interpreted as being a combination of the above meanings, as in a particularly witty or sophisticated type of humor. In other words, there is not a perfect translation of this word and with such a broad range of potential meanings, it is impossible to know exactly what each participant was intending when using this word. For this reason, *gracioso* was kept as its own category in figure 1. With that said, however, the use of *gracioso* seems to be the closest to the types of answers participants used in the English study; it speaks to the idea of formality with an underlying sense of humor or wit. That only five participants used this word could be taken as evidence that perhaps a change is beginning to happen in Spanish similar to the one in English[27].

4.4 Discussion

Results of the corpus study presented in this chapter show a distributional difference in the types of information with which DC is most felicitously used, from that of SC in English. While *said* and *dicho* are both most likely to be used with information that has already been linguistically mentioned or explicitly stated in the discourse, *dicho* is much more commonly used with inferred information, or information that has not been explicitly mentioned in the discourse. While this is very different than in the English

[27] It is also worth noting that there is the possibility that bilingualism and/or contact effects from English are playing a role here. This is worth investigating in future studies

study, it does validate the idea communicated by participants of the social meaning study that Spanish speakers may use this construction to avoid stating something that was already mentioned, and to avoid redundancy[28]. When considering this intuition, we would expect the corpus data to reflect this by having more tokens from the type 4 (inferred) category, and this is attested in the corpus results. While the majority of the corpus tokens still fell into the "mentioned" categories of type 1 (explicitly mentioned, n=107) and type 2 (subtype/type or property/missing property, n=132), 24% of total tokens were type 4, whereas in the English study presented in chapter 2, only 4% of tokens fell into this category. This could arguably show a difference in the functions of this construction across the two languages; according to participants' survey data from task two of the social meaning experiment, Spanish seems to favor it in more formal or "official" types of discourse to avoid redundancy, while in English it is now most commonly used in informal discourses to communicate a subtle sense of humor and wit. While I was able to demonstrate how this shift has occurred diachronically in English, further data would be needed in Spanish to see if a change has taken place, as a genre-balanced data set spanning several decades was not available for this analysis. From the corpus data, however, we can see that this construction is much more widely used in Spanish than in English; the searches resulted in close to 300,000 tokens of DC from the CdE, whereas the COCA searches had fewer than 2000 tokens of SC. Although the CdE is much larger than the COCA (2 billion vs. 560 million), reducing the total of CdE accordingly would still be around 80,000 tokens.

[28] It is also worth noting that the sentences that used DC did not seem (to me, a non-native Spanish speaker) to be sentences that would have used *said* in English.

It could be possible that this difference in frequency is a factor that has contributed to the construction's change in meaning and usage in English, and the seeming lack thereof in Spanish. Returning to the idea of expectations (a la Acton's Violations of Expectations principle), a higher frequency construction would mean that its usage does not violate conversational expectations, thus not flagging significance in a discourse. As discussed in chapter 3, the fact that SC in English is not as commonly used could be contributing to its unique meaning in discourse; its usage violates expectations, flagging significance which English speakers take to mean mock-formality or humor.

These results are also reflected in the social meaning experiment; while the results presented in chapter 3 show that speakers of English interpret SC to have a more humorous meaning than a standard determiner, it is evident that speakers of Spanish do not share this intuition. There was no significant relationship between Spanish participants' use of a 😂 emoji with sentences containing *dicho*, whereas this relationship was significant in the English study. Furthermore, none of the Spanish speaking participants[29] had any reference to humor in their answers to the open ended survey questions, but many of them spoke to its relation to intelligence, formality, or its ability to make a discourse more "interesting". Additionally, other task 2 responses provided data showing both that Spanish speakers were less likely than English speakers to say they encounter this construction in their own social media usage, though English and Spanish speakers self-reported similar estimations for the amount

[29] Other than potentially one or two who used *gracioso*, but this was a problematic case, as described above

of time they spend on social media each week. Spanish speakers were, however, more likely to say that they choose to use *dicho* themselves on a regular basis; English speakers were less likely to report this.

Results from these studies can be argued to support the notion that *said* in English has undergone and is currently undergoing a change from a formal construction to an informal one that carries a unique social meaning, supported by evidence that it has remained relatively unchanged in Spanish, and is used and interpreted as a formal construction which may come across as pretentious in less formal genres of discourse. This study also leaves questions open for both related and unrelated future research. For instance, there is a bit of evidence (namely the use of *gracioso*) that DC in Spanish may be beginning to change in usage and interpretation; a longitudinal study would be useful to see if it has changed thus far and is on a similar trajectory as SC in English. Furthermore, a somewhat unrelated study investigating the reason behind why Spanish speakers wish to avoid redundancy by not repeating identical NPs could provide potential clues as to why *dicho* is more likely to be used with inferred information.

CHAPTER 5

Conclusion

5.1 Summary

The studies in this dissertation have attempted to provide a holistic picture of the history and current behavior of the Said Construction. It has also shown how, despite having shared roots, the usage of these constructions has diverged in English and Spanish, the latter of which has maintained the construction as a more formal one. This project was by no means exhaustive; there are plenty of questions left unanswered and enough future studies for years of *said* research. I hope that it has at the very least served as a conversation starter; this construction has not to my knowledge been brought into the space that is discourse pragmatics, and given its applications to theories of givenness, definiteness and information structure, not to mention its rich history, perceived social meaning, and the presence of sister constructions cross-linguistically, I believe it deserves a closer look. This chapter will summarize the key findings of each chapter, and provide some ideas for continued research on this topic.

5.2 Key Findings

5.2.1 Corpus Study: English

The corpus based analysis presented in chapter 2 presented several key findings. First, this study showed through information status coding that SC favors environments with a linguistically mentioned antecedent. If operating under the assumption that *said* in SC is determiner-like, this points to a stricter, linguistically based model of givenness, as *said* is not likely to occur in sentences that favor models of givenness that include inferences and common world knowledge. Although this is not a rule, it is certainly a preference in the data, with very few exceptions.

The corpus chapter also showed that SC is a dynamic construction, having changed over time in its form. Older tokens of SC are more likely to include an additional determiner, taking on the form *the said N*, while more recent tokens are more likely to lack this additional determiner, appearing as *said N*. This could be related to the parallel finding that SC has over time come to be used in informal genres of discourse more frequently than it traditionally has been, with past tokens more likely to occur in formal types of discourse such as legal documents.

Finally, although balanced present-day corpus data wasn't available for strictly American English, iWeb data from 2017 shows a very high frequency of SC data, especially when compared with data from the Corpus of Contemporary American English. Even if iWeb data could be reduced to be balanced in size with the COCA, there are still more tokens of SC from one year of iWeb data than in the entire time span covered by the COCA. This both supports the notion that SC has become more common (and specifically so in less formal genres of discourse), and also sets up the studies presented in chapter 3 by shifting the question away from SC in general, to SC

applied more specifically to web and social media based discourse, as well as the idea that speakers may choose to use SC to communicate something to the hearer; this could be an unspoken meaning (such as humor) or an idea about the speaker (perhaps that they are wanting to seem more intelligent or formal that they actually are or than that would normally be required given the context of the conversation).

5.2.2 Social meaning: English

The social meaning study provided clear evidence that speakers of English interpret *said* to have a meaning of "mock formality" or humor. In task 1, the social meaning experiment, there was a statistically significant relationship between SC-containing sentences and the use of a *haha* emoji by participants. Sentences using *said* were more likely to be reacted to with 😂 than identical sentences containing a standard determiner, which were more likely to receive a standard 👍 reaction. Furthermore, the survey data from task 2 showed that participants believe SC is most commonly used as a sort of formalism without taking oneself too seriously, or a way to assert authority on a subject with a sort of humorous twist. This finding ties into Acton's Violations of Expectations Principle, which states that an utterance that violates conversational expectations is more likely to be significant in the discourse. I argue that the use of *said*, as opposed to a more standard determiner form like *the* or *that*, violates expectations and is significant because it carries this particular social meaning of "mock formality".

5.2.3 Spanish

The parallel analyses of Spanish showed evidence that this construction, although distributionally similar in the two languages, varies from English in how it is used and interpreted by speakers. The Spanish social meaning experiment did not show a significant effect between the use of *dicho* and participants' use of a *haha* emoji. In this study, identical questions varying only by determiner type (*dicho/a* vs. the standard demonstrative *eso/a*) generally received the same reaction by participants across blocks. I believe this provides evidence that the use of *dicho* does not violate conversational expectations and is therefore not interpreted as having any specific significance in the discourse, specifically in this case social meaning. This notion was also supported by the answers provided by participants in task 2 of this study, many of whom stated that they do not believe that *dicho* means anything specific, and that speakers may only choose to employ it to avoid redundancy. Most stated that it is used mainly in formal discourse. A few participants stated that they believe people use it to sound smarter or more informed than they really are, further supporting the claim that this construction in Spanish is considered more formal than a standard determiner form.

Through these studies, I have shown a potential divergence in usage and interpretation in two constructions with similar roots. While both English and Spanish originally featured these constructions in formal genres of discourse, it has become common in English to use SC in informal discourse to communicate a distinct meaning, whereas Spanish has maintained DC as a more formal construction. These findings lead to questions for future research about how social meanings develop and are perpetuated across discourse communities.

5.3 Study limitations and improvement for future research

5.3.1 English corpus study

The corpus study was limited in that there was not a way to include internet and social media data in the diachronic change portion of the analysis, as these are not recognized genres in the COCA and the data would therefore not be balanced with the other default genres. The iWeb corpus only contains data from 2017, so even taking a 20 million word sample (the number of words added to each genre in COCA each year) and combing it for tokens of SC would only have provided information about that single year, and would not have been useful for the diachronic analysis. If the study would have included the “in the wild” tokens that had been collected over a span of 6 years, it would also not have produced reliable data, as it would not have been balanced against the COCA tokens. To my knowledge, there is not a corpus that covers the time range this study needs, contains web data, and has balanced genres. Furthermore, there is not always a reliable way to look back at web discourse because the very nature of websites allows them to be revised and updated over time. Aside from the website for the 1996 movie Space Jam, I do not know of another website that has remained unchanged since the 1990s. While an internet archiving program or site like Wayback Machine is useful for this type of analysis and will be further employed in the future, its use for this project was limited because it did not have records of many the sites needed to clarify tokens from dead or revised links.

One further note about the corpus study is that the balance of the sample may have been thrown off due to incorrect tagging in the corpus' algorithm that had to be corrected by hand. I noted in chapter 2 that I compiled 54 tokens of SC that were uttered or written before 1950, however many of these were tagged as being from a much later year (generally an academic paper or archive published sometime in the 1990s), but the actual use of SC in the body of the text was directly quoting a legal document or land treaty that was written in the 18th or 19th centuries. While I am confident in the accuracy of my sample because I double checked, read, and coded each entry individually, a new strategy would be needed if a larger sample were to be taken, as many tokens would be labeled as newer than they actually are. This also means that there are fewer actual tokens of SC from the 1990s in the COCA than it seems, because a portion of these tokens are actually much older.

5.3.2 English social meaning experiment

The social media based methodology used in chapter 3 of this experiment could be improved by balancing for gender and race. While my stimuli did include diversity in name and "profile picture", only three of the ten questions participants viewed were male, while the rest were female. A more balanced model or perhaps even a gender neutral model with ambiguous names and profile pictures featuring multiple people per image could go further to account for underlying race or gender biases.

This methodology could also be improved in the future by recruiting even numbers of participants in each age range. While I was able to recruit a minimum and maximum age for participants (in the case of this study, 18-99 years old), I was not able

to further subdivide this into the age groups I created for the analysis. In task 2 of the social meaning study, participants were asked to select their age range: 18-25, 26-35, 36-45, and 46+. I did this so I could see potential patterns between younger and older millennials, as well as participants born before 1980, the conventional start date of the millennial generation. In a pilot run of this study, the recruits happened to be relatively young, with the majority being in the 18-25 category and only three (out of 200) from the 46+ category. When analyzing this pilot data, there was a significant effect between participant age and humorous interpretation, with millennial participants being significantly more likely to use the 😂 reaction for an SC-containing sentence than older participants, who were most likely to use 👍. This result, however, did not hold in the official experiment; there was no significant effect between age and reaction, which could likely explained by the fact that there were more participants in the older age groups. The recruitment platform used for this study, prolific.ac, does not have an option to balance ages within defined groups, but such a feature would allow more reliable data for looking at relationships between reactions and age groups.

5.3.3 Spanish studies

The Spanish studies were the most limited, as I was unable to completely duplicate the English studies due to differences between the COCA and the CdE. In the corpus study, for example, it was not possible to do a diachronic analysis because the web portion of the CdE only covered data from the past 10 (or so) years. This study would be significantly improved by being able to take a larger sample across more genres and a wider timespan, in order to apply the same methodology as was used for

said to track diachronic change. This study would also be improved by being able to take an additional sample from the genre/historical section of the Corpus del Español; this corpus covers the time range necessary for a diachronic analysis, but is only balanced across genres for the 1900s, making it not ideal for this project. Furthermore, this branch of the CdE does not have each entry individually labeled for the specific year as does the COCA, but rather just for the century in which it was uttered. Finally, I was unable to take a random sample that included only data from the 1900s AND from Mexican sources only (as opposed to all Spanish-speaking countries represented in the corpus). These limitations from the genre/historical branch of the CdE made the web/dialects portion of the corpus more usable for this project, even though its time range made a diachronic analysis not possible.

The Spanish social meaning study could be improved upon in the same ways mentioned for the English study: balancing for age group among participants, and gender for stimuli items in the experiment. Furthermore, there were no monolingual Spanish speaking participants available on the recruitment platform, but recruiting completely monolingual speakers for both the English and Spanish studies could result in even more reliable data. Both the corpus and social meaning studies were limited due to the fact that I am not a native speaker, and am less astute to the subtleties of Spanish than a native Spanish speaker would be. While my RA was very helpful in translating stimuli from English to Spanish, she was not available to help with the analysis of experimental data.

5.4 Food for thought

I conclude this project by providing ideas for potential further work on this topic, in hopes that the work presented here has served to spark interest and provide a starting point for future research.

1. Expanding the corpus analysis to include a larger sample of data, especially balancing non-internet data with web and social media tokens.
2. Expanding the crosslinguistic analysis to include other languages with a similar construction. French, for instance, has a similar construction *le dit*.
3. An experiment testing the acquisition of SC in children: Do children, when presented with old and novel discourse entities, interpret uses of SC to mean only the old entity?

The Said Construction is a wide open construction for linguistic analysis across several subfields. The work presented in this dissertation has focused specifically on historical change and social meaning, but much more work can be done syntactically, semantically, and experimentally.

References

Abbott, Barbara. (1993). A pragmatic account of the definiteness effect in existential sentences. *Journal of Pragmatics* 19(39-55).

Abbott, Barbara. (1997). Definiteness and existentials. Language 73:1, 103-108.

Acton, Eric. (2014). Pragmatics and the social meaning of determiners. Doctoral dissertation, Stanford University.

Acton, Eric. (2019). Pragmatics and the social life of the English definite article. *Language* 95.1: 37-65

Acton, Eric and Chris Potts. (2014). That straight talk: Sarah Palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics* 18.1:3-31.

Attardo, S. (2017). *Handbook of language and humor*. New York: Routledge

Birner, Betty J. and Ward, Gregory. (1994). "Uniqueness, Familiarity, and the Definite Article in English." in *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society.* Pp. 93-102.

Bolden, Galina B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation. *Journal of Pragmatics* 41: 974-998

Bowdle, Brian F. and Gregory Ward (1995). Generic Demonstratives. *Proceeds of the Twenty-First Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Historical Issues in Sociolinguistics/Social Issues in Historical Linguistics.* 21(1):32-43.

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change,21*(1), 135-156. doi:10.1017/S0954394509000052

Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and Linguistics Compass*,4/6: 377-389

Chappell, W. (2016). On the social perception of intervocalic /s/ voicing in Costa Rican Spanish. *Language Variation and Change, 28*(3), 357-378.

Cheshire, Jenny. (2005). Syntactic variation and beyond: Gender and social class variation in the use of discourse-new markers. Journal of Sociolinguistics 9(4):479–508.

Comrie, Bernard. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: University of Chicago Press.

Constine, Josh. (2016). Facebook Enhances Everyone's Like With Love, Haha, Wow, Sad, and Angry Buttons. Techcrunch.com.

Davies, Mark. 2008-. *The Corpus of Contemporary American English: 450 Million Words, 1990–Present*. Available online at http://corpus.byu.edu/coca/.

Eckert, Penelope. (1989). Jocks and burnouts: Social categories and identity in the high school. New York: Teachers College

Eckert, Penelope. (2000). Linguistic variation as social practice: The linguistic construction of identity in Belten High, Language in Society. Vol. 27. New York: Blackwell.

Eckert, Penelope. (2012). Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology* 41:87–100.

Freeze, Ray. (1992). Existentials and other locatives. Language 68.553-95.

Frier, Sarah. 2014. Facebook's Second-Quarter Revenue, Profit Tops Estimates.

Bloomberg. <https://www.bloomberg.com/news/articles/2014-07-23/facebook-posts-second-quarter-revenue-profit-topping-estimates>

Goatly, A. (2012). *Meaning and humour*. Cambridge, U.K.: Cambridge University Press.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole and Jerry Morgan, eds., *Syntax and Semantics*, volume 3: Speech Acts, 43–58. New York: Academic Press.

Guéron, Jacqueline. (1980). On the syntax and semantics of PP extraposition. Linguistic Inquiry 11.637-78.

Gundel, Jeanette; Nancy Hedberg; and Ron Zacharaski. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69:274–307.

Haviland, Susan E. and Clark, Herbert H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior* 13:512–521.

Hawkins, John A. (1978). Definiteness and indefiniteness. Atlantic Highlands, NJ: Humanities Press.

Hawkins, John A. (1991). On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(405-442).

Holmback, Heather. (1984). An interpretive solution to the definiteness effect problem. *Linguistic Analysis* 13. 195–215.

Jenkins, Lyle. (1975). The English existential. Tübingen: Niemeyer.

Labov, William. (1963). The social motivation of a sound change. Word 19:273–309.

Lakoff, George. (1987). Women, fire, and dangerous things. Chicago: University of Chicago press.

Leech, Geoffrey N. (1983). Principles of Pragmatics. London and New York: Longman Group Limited.

Li, Charles and Thompson, Sandra. (1976). Subject and topic: a new typology of language. In Li, C. ed. Subject and topic . NY: Academic Press. 457-489.

Lyons, Daniel. (2017). How many people speak Spanish and where is it spoken? Babbel Magazine. <http://www.babbel.com/en/magazine/how-many-people-speak-spanish-and-where-is-it-spoken/

Lyons, John. (1977). Semantics. Cambridge: Cambridge University Press.

Lyons, John. (1999). Definiteness. Cambridge: Cambridge University Press.

Martínez, Ignacio M. Palacios. (2011). "*I might, I might go I mean it depends on money things and stuff*". A preliminary analysis of general extenders in British teenagers' discourse. *Journal of Pragmatics* 43: 2453-2470

Mata, Rodolfo. (2016). "Traversing the wall: A Study of Language Contact among Heritage and Immigrant Speakers of Spanish in the Tijuana-San Diego Border Area." Doctoral dissertation, University of California, San Diego.

Mendoza-Denton, Norma. (2008). Homegirls: language and cultural practice among Latina youth gangs. Malden, MA: Blackwell Pub.

Milsark, Gary. (1977). Toward an explanation of certain peculiarities of the existential construction in English. Linguistic Analysis 3, 1-30.

Moreau, Elise. (2018). How to use Facebook Reactions. Lifewire.com: https://www.lifewire.com/how-to-use-facebook-reactions-3894307

Norberg, Dag. (1980). Latin at the End of the Imperical Age. *Manuel pratique de latin medieval*. http://www.orbilat.com/Languages/Latin_Medieval/Dag_Norberg/01.html

Norrick, Neal R. (2009). Interjections as pragmatic markers. Journal of Pragmatics 41: 866-891

OED online. (2019). said, adj. and n. Oxford University Press. https://www-oed-com.proxy.lib.umich.edu/view/Entry/169800?isAdvanced=false&result=1&rskey=KkMm5v& (accessed October 20, 2019).

Prince, Ellen F. (1988). The discourse functions of Yiddish expletive -*es* + subject-postposing. IPrA Papers in Pragmatics 2. 176-194.

Prince, Ellen F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S.Thompson and W. Mann, eds., Discourse Description: Diverse Analyses of a Fundraising Text. Amsterdam/Philadelphia: John Benjamins. 295–325.

Rando, Emily N. and Donna Jo Napoli. 1978.Definites in there-sentences. Language 54.300-13.

Rosch, E., C.B. Mervis, W. Gray, D. Johnson, & P. Boyes-Braem. 1976. Basic objects in natural categories. Cognitive Psychology 8:382-439.

“Said.” *grammarist.com*. Grammarist 2014. Web. 30 August 2019.

“Said”. *Macmillandictionary.com*. Macmillan 2019. Web. 30 August 2019. https://www.macmillandictionary.com/us/dictionary/american/said_1

"Said". *Merriam-webster.com*. Merriam Webster, 2018. Web. 4 May 2018. http://www.merriam-webster.com/dictionary/said

Schwarzschild, Roger. (1999). GIVENNESS, AVOIDF, and other constraints on the placement of accent. Natural language semantics, 7(2), 141-177.

Stevers, Alicia. (2014). Syntactic roles and functions of the Said Construction. Master's thesis, San Diego State University.

Stevers, Alicia. (2017). Givenness and the Said Construction. Qualifying Research Paper, University of Michigan.

Stevers, Alicia. (2018). Social meaning and the Said Construction. Under review, *Journal of Pragmatics.*

Stubbe, Maria & Janet Holms. (1995). *You know*, *eh*, and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication* 15(1): 63-88

Tiersma, Peter M. (1999). *Legal Language.* Chicago, IL: University of Chicago Press.

Trudgill, Peter. (1983). On Dialect: Social and Geographical Perspectives. NY: NYU Press (151)

Ward, Gregory and Betty J. Birner. (1995). "Definiteness and the English Existential," in *Language* 71:722-742.

APPENDIX A

Corpus of US Supreme Court Opinions
Word and Text Breakdown (Davies 2008)

Decade	# words	# texts
1790s	603,835	334
1800s	428,193	176
1810s	1,032,229	362
1820s	1,607,874	376
1830s	2,566,250	462
1840s	2,263,626	395
1850s	3,337,263	875
1860s	2,920,840	974
1870s	6,035,231	1947
1880s	7,789,600	2487
1890s	10,096,146	2480
1900s	7,266,439	1998
1910s	5,925,516	2397
1920s	4,541,174	2021
1930s	4,804,810	1769
1940s	6,662,677	1485
1950s	4,669,109	1342
1960s	7,674,366	2674
1970s	11,968,472	2204
1980s	14,383,294	2185

1990s	9,166,448	1130
2000s	7,542,515	1025
2010s	6,138,407	1027
TOTAL	129,424,314	32,125

APPENDIX B

Social meaning experiment:
Stimuli, English study

Said	Standard Det
A 10 lb baby doesn't sound heavy...unless said baby isn't content & won't nap unless you are holding them & standing up.	A 10 lb baby doesn't sound heavy...unless that baby isn't content & won't nap unless you are holding them & standing up.
Having a kitty with a cold who wants to snuggle is adorable. Getting sneezed on by said kitty is less so.	Having a kitty with a cold who wants to snuggle is adorable. Getting sneezed on by that kitty is less so.
So I overhear my daughter playing a game with a horse… what has she named said horse? Shadowfax. She is the coolest kid.	So I overhear AJ playing a game with a horse… what has she named the horse? Shadowfax. She is the coolest kid.
Well, tonight was the night I made my husband make the Panera people re-make me a sandwich even though they no longer have the ingredients to make said sandwich.	Well, tonight was the night I made my husband make the Panera people re-make me a sandwich even though they no longer have the ingredients to make the sandwich.
We're adults, we bought a house, we also may or may not be currently playing hide and seek in said house	We're adults, we bought a house, we also may or may not be currently playing hide and seek in that house
When you're babywearing your newborn while eating a doughnut and you look down to find said newborn covered in crumbs…	When you're babywearing your newborn while eating a doughnut and you look down to find the newborn covered in crumbs…
So, the day before you start teaching a class about a particular subfield is probably the right time to decide you're totally wrong about the underpinnings of	So, the day before you start teaching a class about a particular subfield is probably the right time to decide you're totally wrong about the underpinnings of

said subfield, right?	that subfield, right?
I'M IN NEED OF A 55 GALLON DRUM… for our men's desert trip. Yes, it will be awesome. Yes, I will send you a photo if you give me said drum.	I'M IN NEED OF A 55 GALLON DRUM… for our men's desert trip. Yes, it will be awesome. Yes, I will send you a photo if you give me the drum.
So... in theory I like sweet potatoes, but said theory includes loads of brown sugar, butter, and marshmallows.	So... in theory I like sweet potatoes, but the theory includes loads of brown sugar, butter, and marshmallows.
Hypothetically, if someone really misses playing a real piano, who do I know that has one in their home and wouldn't mind letting said person come by to play it every once in a while? Asking for a friend.	Hypothetically, if someone really misses playing a real piano, who do I know that has one in their home and wouldn't mind letting that person come by to play it every once in a while? Asking for a friend.

APPENDIX C

Social meaning experiment:
Stimuli, Spanish study

Dicho	Standard Det
Un bebé de 5 kilos no suena pesado... a menos que dicho bebé no esté contento y no tome siesta y que lo estés cargando mientras estás parado.	Un bebé de 5 kilos no suena pesado... a menos que ese bebé no esté contento y no tome siesta y que lo estés cargando mientras estás parado.
Tener un gatito con resfriado que quiere acurrucarse es adorable. Que te estornude dicho gatito no lo es.	Tener un gatito con resfriado que quiere acurrucarse es adorable. Que te estornude ese gatito no lo es.
Así que escuche a mi hija jugar un juego que tiene que ver con un caballo...¿como le ha llamado a dicho caballo? Galante. Ella es una niña genial.	Así que escuche a mi hija jugar un juego que tiene que ver con un caballo...¿como le ha llamado a ese caballo? Galante. Ella es una niña genial.
Bueno, hoy fue la noche en la que hice que mi esposo hiciera que gente del restaurante volviera a hacerme mi sándwich a pesar de no tener los ingredientes para hacer dicho sándwich.	Bueno, hoy fue la noche en la que hice que mi esposo hiciera que gente del restaurante volviera a hacerme mi sándwich a pesar de no tener los ingredientes para hacer ese sándwich.
Somos adultos, compramos una casa y pudiera ser o no que ahora mismo estemos jugando a las escondidillas en dicha casa.	Somos adultos, compramos una casa y pudiera ser o no que ahora mismo estemos jugando a las escondidillas en esa casa.
Cuando cargas a tu recién nacido mientras comes una dona y miras hacia bajo para encontrar a dicho recién nacido cubierto en migajas de pan.	Cuando cargas a tu recién nacido mientras comes una dona y miras hacia bajo para encontrar a ese recién nacido cubierto en migajas de pan.
Así que el día antes de comenzar a enseñar una clase acerca de un tema es	Así que el día antes de comenzar a enseñar una clase acerca de un tema es

probablemente el mejor momento de decidir si estás totalmente mal acerca de los fundamentos de dicho tema.	probablemente el mejor momento de decidir si estás totalmente mal acerca de los fundamentos de ese tema.
Necesito un barril de 200 litros... para nuestro viaje de hombres al desierto. Si, va a ser increíble. Si, te mandaré una foto si me das dicho barril.	Necesito un barril de 200 litros... para nuestro viaje de hombres al desierto. Si, va a ser increíble. Si, te mandaré una foto si me das ese barril.
En teoría me encanta el plátano macho, pero dicha teoría include que se fían, mucha azúcar y crema.	En teoría me encanta el plátano macho, pero esa teoría include que se fían, mucha azúcar y crema.
Hipotéticamente, si alguien extraña tocar un piano de verdad, a quien conozco que tenga uno en su casa y a quien no le importaría dejar a dicha persona ir a tocarlo de vez en cuando? Preguntando por una amiga.	Hipotéticamente, si alguien extraña tocar un piano de verdad, a quien conozco que tenga uno en su casa y a quien no le importaría dejar a esa persona ir a tocarlo de vez en cuando? Preguntando por una amiga.

APPENDIX D

Social meaning experiment:
Example of experiment layout, English Study

Jamie W.

A 10 lb baby doesn't sound heavy...unless said baby isn't content & won't nap unless you are holding them & standing up.

Appendix E

Social meaning experiment:
Example of experiment layout, Spanish Study

Jaime W.

Un bebé de 5 kilos no suena pesado... a menos que dicho bebé no esté contento y no tome siesta y que lo estés cargando mientras estás parado.

APPENDIX F

Social Meaning Experiment:
Task 2 Survey, English

1. Select your age range

18-25 26-35 36-45 46+

2. How many hours a week would you estimate you spend on social media?

0-2 3-5 6-8 9+

3. Have you encountered sentences like the one below, that use the word “said” in this way?

“I got a coffee at Starbucks, but then I spilled **said coffee** on my new sweater.”

Yes No

4. Do you ever use “said” in this way? How often?

Never Rarely Sometimes Regularly Frequently

5. (Open ended question)

Why do you think people use “said” in this way? Does it have a purpose?

6. (Open ended question)
Complete the following sentence:
I think people use “said” when they want to seem _____.
(Answer may be more than one word if needed)

7. (Open ended question)
How do you think its meaning differs from other words that could take its place, like “the” or “that”?

For example, “I got a coffee at Starbucks, but then I spilled that coffee on my new sweater.”

APPENDIX G:
Social Meaning Experiment
Task 2 Survey, Spanish

1. Por favor seleccione su rango de edad

18-25 26-35 36-45 46+

2. ¿Cuántas horas por semana estimaría qué pasa en la red social? (Incluya todas las redes sociales: Facebook, Instagram, Snapchat, Tumblr, Twitter, Myspace, WhatsApp, etc)

0-2 3-5 6-8 9+

3. Lea las siguientes oraciones. En su uso normal dentro de las redes sociales, ¿Se ha encontrado con oraciones que utilizan la palabra “dicho” de este modo?

“Compré un café en Starbucks, pero enseguida derramé **dicho** café en mi suéter nuevo.”
“Ha estado a 32 grados por una semana entera y basta ya de **dicho** calor.”

Sí No No sé

4. ¿Alguna vez ha utilizado “dicho” de esta manera? ¿Qué tan seguido?

Nunca Rara Vez Algunas veces Regularmente Frequentemente

5. ¿Por que piensa que la gente usa “dicho” de esta manera? ¿Tendrá un propósito? (Escriba su respuesta en el espacio siguiente)

6. Complete la siguiente oración:

Yo pienso que la gente usa “dicho” cuando quieren parecer _____.

(Su respuesta puede ser más de una palabra si es necesario)

7. ¿Cómo cree que “dicho” sea distinto en su significado entre otras palabras tales como “el” o “la” o “ese” o “eso”? Por ejemplo, “Compré un café en Starbucks, pero

enseguida derramé ese café en mi suéter nuevo.”

APPENDIX H

Social meaning experiment:
Task 2 question 5, English study

Why do you think people use "said" in this way? Does it have a purpose? (Type your response in the space below)
to be funny
To indicate or refer back to a previous statement. I assume they are doing it to shorten the sentence and conserve characters for the rest of the message especially if there is a restricted character count like Twitter.
I think they are trying to mimic the sound of a cheeky, personal diary and serves no other function.
Not sure most likely just a habit picked up from someone close to them growing up.
To reference something previously mentioned
To indicate you are speaking about the same thing as before.
I think people think it makes them sound funny.
To emphasize the sentence.
To make their sentence sound more interesting, to avoid using "the" or "it"
It's funnier than "the" baby or "the donut." It reads as more emphatic and adds an air of bemused frustration.
To emphasize that one particular said thing
Explaining the past experience
Emphasis and yes
To make themselves sound/appear more intelligent
They think they sound wiser and more intelligent.
To add emphasis on on the action that is occurring.

To make themselves sound more educated.
It is any easy way to add a formality to a simple statement in a way that conveys humor.
They are silly!
I think it's just the evolution of words. New contexts for things.
Bringing attention to the fact that the coffee you are referencing is a coffee previously used in an example
It gives emphasis to the item causing discord.
It's to point out what the object of the sentence is without redundancy.
Usually they're explaining something and think it makes them sound smarter.
It doesn't have a purpose used that way. I think people use it because they think it makes them sound smarter than they are.
indicating that they are talking about the same exact noun
It has a purpose but it feels like more of an outdated purpose. I feel like I've encountered it more in the past than I do now.
To indicate who/what they're talking about
To emphasize that the same object is being talked about
it like a past tense use of that.
to shorten up aforementioned.
It can be used to distinguish what you're talking about, instead of using a more generic 'it'
I think its a funny way to refer to common social situations.
To me, it just makes some sentences sound more humorous
I think people use that word because their have poor grammar and sentence structure skills. It has no purpose other than making you look undereducated.
Different way to state something, possibly try to sound smart
to emphasize a situation
I'm not sure if it has a specific purpose, but it's another way of saying "that thing I just mentioned"
Wierd

to emphasize something that came before.
I think using "said" in the manner highlighted appears more grammatically correct or proper.
It gives more emphasis on the word right after said.
People read other posts that included "said" and started to include it in their own posts
I think people use "said" in this way to make themselves sound smarter.
My first reaction is because most people are idiots. To emphasize the following word.
I think it can put an emphasis on what is being talked about.
Sound smart
To sound educated.
It doesn't have a good purpose from my perspective. I think it's a tacky style of writing.
Picking up language used in popular jokes
It is a way to casually emphasize a target noun that was already brought up. Usually for humorous effect.
yes
To refer to something previously mentioned in a statement
I really have no idea, it seems strange and a little childish.
It adds emphasis to the earlier part of one's story
They are trying to sound smarter
Replaces 'that specific [noun]' or having to re-refer to something in conversation
Emphasis on the object
provides emphasis
It's used as an easier-to-say synonym to "aforementioned" and imparts a specific tone to the words.
It sounds complex and intelligent
To connect it to what they said previously
saying the said coffee seems to me like saying: the (recently referenced) coffee

To emphasize or focus on a specific thing
It's a way to provide emphasis on a particular object that was announced earlier in the post. It doesn't serve any true purpose, it's a piece of social media speech that has evolved.
Being sarcastic , funny and or playful
To emphasize what happened.
It's an easier way to say "aforementioned"
This use of "said" helps to clarify what a person is talking about, by referring back to something previously said. Hence, "said."
To express their unique grammatical tastes? It reads differently and makes you pay more attention while reading perhaps? I'm not really sure. Maybe some people have spoken like that for years and speak in said fashion often so they reflect that in their social media posts.
They can't think of anything else.
It is indicating that you are referring to an object/person/etc. that had already been referenced
It just seems like a replacement for using "this" or "that". So I guess its purpose would be to add variation to the way you say something. Or to also refer back to something previously mentioned, but when it was just mentioned, it seems a bit redundant.
Perhaps they have heard or seen others use that wording.
Probably because they've seen other people use "said" in that way. I don't know what popularized that usage in the first place.
I don't really think it has a purpose. It's not necessary.
Just to add emphasis on the particular item that was altered in some manner. I use it more or less when I am trying to be coy, as if the event were hypothetical and did not actually occur when it really had.
I believe that it elevates the object above its normal position in conversation
Honestly beyond it's original use, I think now people use it to sound more formal in a comedic way.
Just trying to be fancy
It alludes to a hypothetical thing, or an alleged thing.

They hear other people say it and it becomes a trend
To describe a purpose or item in a situation
I don't know, I wish they wouldn't. It gives me a more negative impression for some reason. I really don't think it serves a purpose, maybe they think they like the way it sounds.
To me it almost means supposed or so called. Not so much like factual.
i dont really know why they do that.
I'm not really sure. Maybe to try to be funny, to sound kind of serious about something silly?
Another word for the
I actually have no clue, Me and my friends never use this term ever. lol
It's just a different way of referring to something previously referenced.
TO STATE IT IS THERE TO MAKE POINT.
Sounds pretentious on social media
I am not sure. Maybe people think they are being fancy.
It signals a joke usually, but like a witty or intellectual one
make them sound more smart
It feels like people trying to be copy cat
Maybe to make a point or put emphases on the object
I think they think they are being cute, but in reality, not.
its referring to the aforementioned noun.
mostly for emphasis
To distance themselves from the subject yet feel attached.
I think most of the time people use it because they think it's the more proper way to phase their sentence/question.
It sounds fancy and special but doesn't really mean anything other than a waste of time and space.
No clue
not really. it makes them sound brighter.
Doesn’t seem to have much of a purpose. It may grammatically but in every day use

it seems a little over the top
to reflexively indicate the object
it draws attention do a different part of the sentence and makes you sound more intelligent.
To sound more intelligent and inject some length to their short nonsense.
It's a hipster-y thing, I would say. It's fun to say things in different ways.
It's the way that they talk
No idea
Emphasis on a particular word or subject
"said" refers to the item you previously mentioned
To reference the previously mentioned noun in a way that sounds somewhat academic
they think it makes them sound funny
yes
Maybe to exphasize what you’re talking about
the truth is that it is something that is used without thinking
It's a method to emphasize the word that follows after "said".
I think they're trying to place more emphasis on whatever it is they're saying
they think it makes the sentence sound more clever/funny
weird
Mistyping
To make their point clear. It does sound cute.
ease of use
I think it makes things clearer for them.
It adds emphasis to the particular item mentioned.
People always want to response for establishing his/her existance.
It is reserved for a formal, businees-like or legal context.
They think it sounds cool, there isn't a purpose though.
To specify a certain item.

To elicit a specific tone.
To sound more intelligent
I think people use it to reference something they have already alluded to previously, so as to avoid being redundant.
something
I usually see it in cutesy internet-joke speak circumstances, I think it's often used like a faux-formality meant to be funny.
It does not really fit with in with the other social media posts. If someone where to use it I would assume it would be part of a joke.
It seems different than the grammar that is normally used on social media
For attention I guess.
It totally has a purpose in conversation. It was mostly made by black femmes and we should respect where that word came from
I am not equipped to explain the linguistic nuances of this to you
I do not have a clue.
just explaining the situation.
trying to be conversational
I don't know if it has a purpose, but it seems to bring extra emphasis to the object.
I guess it's used in place of a pronoun, its a very lawyerly word.
To sound pretentious
as a reference and emphasis on an object they want to make attention towards to
It refers to something you've identified earlier in the sentence, I believe
To refer to the aforementioned subject
For emphasis I suppose.
it is right way
not sure
Now that I think about it, it doesn't really have a purpose in casual conversation.
I think it is to be sarcastic or cynical.
to refer to the thing they spoke of previously in the comment
To sound fancy.

I think they are trying to be funny.
doesn't have any purpose or meaning to me
Using it almost as a way of saying "aforementioned"
To avoid repeating the aforementioned noun
It adds a tone of humor to the sentence
Specifies object in a formal way
no
To emphasize the object of said
to put extra focus on the noun
yes
To exaggerate more
yes, it emphasizes the action or object in the sentence
makes them sound more sophisticated
it's just the way some people write; to me, it sounds better than using 'it' repeatedly
to personify objects
I'm not sure.
It just sounds cool.
It is to refer back to the main item of the topic previously stated.
yes
To emphasize a point or its humor/seriousness
pretty much ignorance of the English language. Very occassionally used to draw attention to the post
To be funny
is the past
It's referring to the previously stated item; a convenient way to reference what you had 'said' before
Sound witty
so they can sound smart. i dont think it has a real purpose
To make themselves sound smarter

Because it makes sense in context
I'm not sure. It comes off as awkward for me. I don't see a purpose.
I’m not sure
Unsure, I think it's more common in funny posts.
I think people use it to make themselves sound educated
It seems perfectly acceptable; just a different way to say something, not a word I would ordinarily use!
They used 'said' in that way to reference the object from earlier in their sentence. It could just have easily been replaced with "I spilled that coffee" or "I'm over this heat", but its still acceptable to have "said" there instead.
Said means the same thing as that in that (or said) context.
it's self-referential, but with specific emphasis, typically for comedic effect
I think people use it primarily as a replacement for "this", "the", "aforementioned", etc. It's purpose in this context is to be an identifier
Because it's a correct way to use the word. It's a good way to refer to something previously mentioned.
humorous way to refer to something already mentioned
It’s accucasatory
To add emphasis to what they are talking about
Though it was normal grammar...
To sound cool. No purpose
They´re just using a different word(said), in a phrase, that better expresses their thought, nothing special.

APPENDIX I

Social meaning experiment:
Task 2 question 6, English study

Complete the following sentence: **I think people use "said" when they want to seem _____.**
funny
Smarter
cute or witty
Witty
smarter
funny
funny
angry
more formal
Bemused and vaguely irked.
Irritated
funny
more intelligent
smarter
fancy
educated.
educated, but don't take themselves too seriously.
basic
hip
Pedantically specific

Justified in their anger
some people might think "fancy" and i've seen people use it that way, when they're seriously or sarcastically trying to come across as refined, but i don't think there's always an ulterior motive.
smart.
intelligent
smart
Sophisticated?
clear
dramatic
formal
snarky
more intelligent
Smart. Witty.
whimsical
unique and different.
intellegent
funny
clear about what they're saying
Smart
funny
mature
intelligent
well read
smart.
dramatic
smart
educated
educated

redundant
ironic
emphatic
intellectual
fool
funny, intelligent
smarter
Humorous but not uptight
smart
quirky
well-spoken
educated
smarter
Clear
normal?
more specific
Playful
interesting
precise
clearer and more formal.
different
smart
more intelligent
educated. Like they are trying to sound a little more fancy by spicing up their sentence.
educated or trendy.
intelligent and funny
intelligent
coy

elegant in speech
tongue-in-cheek, formal
intelligent
fancy
quirky
Smart or educated
intelligent
Like they can add words to their sentences?
witty
serious but silly.
Intelligent
sophisticated
educated
Smart
fancy
Smart and funny
smart
Smart
like a lawyer. (making a point)
smarter
pretentious
annoyed
knowledgable
More proper in their wording.
special
smart
Smarter
snarky
intelligent

more intellectual
Silly
intelligent
Intellectual
Smart
like they're emphasizing a previously mentioned item
cynical
funny
Like they are emphasizing something...
different
Intelligent
more educated
smart or funny
weird and extra
like they know what they are talking about.
sarcastic
smart
smart
angry
right
formal
sophisticated.
annoyed
sarcastic
more intelligent
clear
Funny
Funny
intelligent

attention
funny and clear
stupid.
matter of fact
humorous
clever
silly-formal
pretentious, smart, affected
focused
More intelligent or funny
Intelligent
clever.
smarter
sarcastic
smart
fancy
funny
self-important
like they have a more expansive vocabulary
intelligent
Snarky
formal, in an ironic way
smarte
Emphatic
better person
overreact
funny-serious
worldly, sophisticated
articulate

funny
important
cool
academically inclined
better person
intelligent
silly
funny
past argument
It's a little pretentious, but it's common enough that I don't notice it very much. I may also assume someone is trying to be funny.
witty
cultured or refined
smarter
normal
smarter than others
Normal
funny
Educated
smarter
I think people use "Said" when they want to seem normal, regular, or average.
emotional
funny
Well-spoken
grammatical
wry
Smart
smart
Different

different.

APPENDIX J

Social meaning experiment:
Task 2 question 7, English study

How do you think "said" differs in meaning from other words that could take its place, like "the" or "that"? For example, "I got a coffee at Starbucks, but then I spilled that coffee on my new sweater."
it's funnier
It doesn't differ in meaning. It is a different way to state something that may or may not project an image about the author.
it's douchier
I don't think it differs in meaning when used like this.
It sounds more grammatically correct to say said. I dont think it changes the meaning at all.
It's funnier/snarkier because it looks like you're feigning intellectualism
I think its easy to infer the meaning of said from context. Both could be used.
Said adds more emphasis to the sentence than that would.
Theoretically it should have the same meaning. To me, saying "said coffee" vs. "that coffee" makes the person sound more irritated
It's a little pedantic, I guess. "That" is less academic in the above example, but still funny. Funnier than "the" would have been.
More factual less emotional
that
it's more memorable
I don't think it differs very much other than we don't use it as often
no difference
Using said adds emphasis/definition to the action or object that is being spoken of.
said is speaking as if it has feelings or is a prominent object. Generally, people use this term in the 3rd party again to make themselves sound educated. When in reality

it is silly.
It sounds "fancy"
Well yes..
I think it's similar, but it's a little more specific.
really doesn't
Using "said" conveys a sense of urgency.
it doesn't.
It's less common so it stands out more.
It doesn't differ, just a different way of saying it.
it doesn't really
I'm...not sure. It has a more...high brow feeling to it? It also feels a little more specific?
it sounds more extra
It is more unusual, people might feel more unique for saying "said"
i think said sounds better
it feels less informal
I don't think it really differs in meaning much
I don't think it differs.
It has a nicer ring to it, but overall it means the same thing to me
No, it isn't different. It's just a way to make your sentence sound unique.
No difference in meaning
it's not as emphazied
I think it draws attention again to whatever you're talking about - "spilled that coffee" sounds unnecessarily wordy, like it should just be "spilled it", but "spilled said coffee" sounds better, and sort of like a reminder: "it was COFFEE! and I spilled it!"
None
it seems to change the meaning to more of a light hearted tone
I think said refers to that very specific item one is mentioning in their sentence.
Said just adds more emphasis.
I think it emphasizes the word coffee compared to "the" or "that"
Said is more pretentious.

I think said refers to the act of speaking and should not be used in this context.
It’s basically the same thing being said but it’s a little different because I think saying “said” makes you put emphasis on the word said.
makes you think more
I don't know.
I don't think there's a huge difference. It's just a more obscure way of being redundant.
calling attention to humorous effect
I'm not sure it really does, that much. It's mostly used for a connotation difference, not a semantic one.
he
it sounds more sophisticated
I think using the term "said" when it's not necessary makes people look silly.
"Said" has more emphasis on what was previously said and flows better than "that"
It is referring the the sweater that they already mentioned earlier
Not necessarily different in meaning, but sentence 'sounds' better with said. If using 'that' it would sound better as two sentences.
doesn't feel as genuine
it puts the focus on the next word in a stronger way
In the example, using "that" implies that there were other coffees in the person's anecdote, "said" just carries a different context in everyday speech.
It avoids repetition
no different in meaning
Said can separate which coffee, because it is the coffee that was previously spoken about.
Doesn't really differ. Both said and that emphasize a specific coffee.
Said sounds more informal, more like how you would speak to a person instead of just writing the words.
Said makes the phrase more interesting
It doesn't seem as interesting.
Only in the sense that it specifically points out that it was mentioned before in the conversation

"Said" is more formal and less natural than "the" or "that." It makes someone sound like they're trying to appear smarter.
I'm not sure it does in that context. It's early and I am not fully awake yet to be fair.
a lot
In this context, it doesn't really differ.
It doesn't seem like there is any difference in meaning. It does give a different feel to the sentence though.
It has a more formal sound but means the same.
There's no difference in meaning, but people might find "said" a more interesting word choice because it's used less often than other words that could take its place.
I don't think it, technically, differs. It mean the same thing used in this context.
It has a totally different feel while having the same intent.
it sounds more sophisticated.
I think it is used more consciously, and therefore carries more emphasis.
It does stop repeated use a the main word.
said makes you go back to the beginning of the sentence.
I don't think it differs in meaning at all
they are using it to refer to the coffee they already mentioned, it doesn't really differ in meaning from that in this case.
They’re referencing what they just talked about, it seems unnecessary.
it is more generic to use that. people are afraid to be generic or basic.
It refers to it more personified than like an object.
None
I think they emphazise the coffee more with "said" like they were disgusted in it happening.
The meaning is certainly similar although the use of "said" seems to dictate a more sophisticated use of language.
NO DIFFERENCE
It sounds more natural in day to day conversation
Maybe it is more hip or cool.
I think it implies an ironic tone
sounds more smart

No difference except Said just sounds showoffy
not really different, just points out or directs attention to the object
It's awkward and seem inappropriate
I think they can be used interchangeably
"said" brings more attention to the item and what happened
I don't believe there is a different meaning.
I think they mostly differ due to our perceptions of the words themselves, and how often/common it is for us to see words used in certain ways or contexts.
Said is more pretentious.
it doesn't mean anything different
it sounds weird
Seems more formal
folks trying to be cute. cunts and dicks.
it's more detailed.
It adds some flair to the sentence. Compared to heavy usage of "that" or "like."
It has a sillier connotation
IT sounds more uptight
Said sounds better
Brings more emphasis to the word
I think "said" sounds a little more formal, which can come across as sarcastic and humorous when used the right way
The meaning is the same, "said" just spices up the sentence a bit.
it doesnt
yes
Just to emphasize that that particular and rude coffee spilled on you. I feel like said gives more life to that particular object.
the meaning is the same, they are just different styles
Using the word "said" emphasizes the coffee that was mentioned in the beginning of the sentence.
not really much of a difference in meaning, I think "that" sounds better
I think it sounds more unique and kind of adds a flair to the sentence

normal
Its not as confusing as saying "said"
Said points clearly to the item being discussed.
sounds better
the word still does what "the" or "that" would have done in its place. The word is simply as unnecessary as the word "aforementioned" would be
It adds more emphasis
the
It underscores the main subject of context.
That sounds more traditional.
It's more specific
It sounds more proper
It sounds more intelligent
it sounds a little more sophisticated.
none
It sounds more formal/technical, which I guess is supposed to be juxtaposed with something lighter or more humorous.
It is not really the way people talk in the south. We would understand you but it just sounds odd.
It means the same this, perhaps it is a little more specific to say said
I really have no idea.
all the
it's just a thing people say bro it flows better imo
I think is some kind of "jerga" or "group talking".
its just another way to say it, I dont think it changes the meaning.
more pronounced, noticeable
It means the same thing, but sounds maybe a bit more erudite
As I said - it is something lawyers use so I think people are trying to sound mock-formal
less common, sounds pretentious
said implies attention rather than reference

It sounds more sophisticated. Saying "that" seems redundant
It's a less commonly used reframe. Comes off as more neat and tidy in my opinion.
puts more emphasis on the word.
very fine things.
its the same
I don't think it differs that much. Possibly to emphasis it?
I guess if people are trying to seem more intelligent by using more complex words. "Said" is also less personal. Like something you don't care for or can easily push aside?
i dont think it really does
It doesn't.
Said highlights the object more.
it makes people sound pretentious
sounds more eloquent
I think when someone uses "said" instead of "that" it could mean any adjectives that went along with the noun. For example, if they actually described the coffee with adjectives, they wouldn't need to use all those adjectives the second time.
It makes it more anecdotal
more formal
no
It doesn't
it doesn't differ
he
By the way it is written
using said is less common and so more attention getting and often more droll
the example sounds redundant
no difference
it's more of a humorous tone, rather than factual
They're not that different. "Said" means aforementioned and "that" can mean the previously mentioned thing.
brings more attention to it and yourself.
It is more old-fashioned.

he
It makes the sentence sound less awkward and emphasizes the word
Said indicates a quote...the others do not
Not as funny
i don't know
I think that 'said' flows better than 'that' in the example sentence. If I were to type the example sentence, I would use "I got a coffee at Starbucks, and then spilled it on my new sweater."
It sounds more creative
no difference
said more explicitly and definitively refers to an object already referred to in the past
makes it simpler
It still feels a bit weird. Saying My is such a simpler way to word it.
More explicit, it’s a bit of a shorthand
I think it's very similar.
It sounds more formal to use the word "said" than another word.
I think it gets your attention because it is not used very often!
Using 'said' instead of 'that' doesn't have a huge difference in meaning, it's just expanding the word variety used.
It’s accusatory
Said makes the person sound more emotional.
"Said" has a distinct temporal nature to it, meaning "aforementioned." "That," on the other hand, can have a physical component to it (e.g. "that tree (over there)" vs. "this tree"), but may also derive temporal nature from the context of the sentence, as in the coffee example above. "The" lacks the specificity of either "said" or "that," though it is clear from the context of the sentence what "the coffee" refers to.
It seems to formalize an otherwise unformal sounding sentence
They mean the same thing in that context.
"said" is funny, the others are not
It’s less threatening
"said" sounds better
Said sounds better.

I don’t think it is much different other than a person wanting to stand out or speak a different way than others
I do not think there´s any difference between the words. It is just a preference of one word over another.

APPENDIX K:

Social meaning experiment:
Task 2 question 5, Spanish study

¿Por que piensa que la gente usa “dicho” de esta manera? ¿Tendrá un propósito?(Escriba su respuesta en el espacio siguiente)
Para parecer inteligente
pues es una manera de especificar que se refiere a algo ya anteriormente mencionado
porque tiene un amplio vocabulario
Por costumbre.
Porque es correcto, pero no muy común
La gente utiliza algunas palabras de una manera extraña, ya que no saben el uso correcto de estas palabras, sobre todo en redes sociales.
Para no repetir palabras y tratar de sonar mas interesante
porque es mas comun, es mas simple
para hacer referencia a la persona u objeto sin tener que repetir el nombre o descripción
Para no ser repetitivo
para hacer énfasis en el objeto
lo usan para evitar mal entendidos y hacer énfasis a lo que se refieren
tal vez es algun tipo de persona que no le gusta dejar las cosas a interpretacion de la gente
una manera formal de hablar
Se refiere al sujeto mismo de la oración, para no perder el contexto.
Dicho suena más formal
Parecido a "tal" o "ese".

Para referirse a un objeto antes mencionado
No es de uso común. Suena forzado, como si fuera traducido.
es por mencionar que es, como hablar de algo en pasado en tiempo presente, es una usansa del español que no existe en el ingles
Para demostrar propiedad al hablar/escribir.
No repetir el sustantivo al que se es referido en la oración
Es una manera formal de evitar la repetición de una palabra. Tal vez demasiado formal para un post en una red social.
Porque saben el significado de la palabra y quieren darle una variación a la forma en que escriben sus mensajes.
sustituir palabras
Afirmación. Mala Dicción.
Da a entender de manera rapida
Es propia mía. Sólo me.incluye a mi
para dar contexto a lo ocurrido
Para señalar el objeto del que se habla, de manera formal.
Probablemente para denotar un cierto grado de dominación hacia la situación
Es una palabra común del español.
Parecer mas culta
tendencia, moda... o porque les gusta, pero es raro ver a cada rato: "dicho"
para hacer una connotación de lo que están hablando
Para hacer referencia al sujeto de una oración pasada y no tener que repetir información
para hacer enfasis
Para escribir un poco más formal al referenciar el objeto del que estén hablando.
Para no repetir la misma palabra
Por querer ser especiales.
Representa cierta formalidad en la frase
para destacar algo no habitual
Indicar de lo que se habla sin repetir lo mismo.
No realmente, seria solamente para darle un enfoque mas formal

para utilizar un vocabulario más amplio
Para aclarar
Suena bien
No creo
Por formalidad
Suena mas a una manera pretenciosa y muy desarrollada de hablar en redes sociales.
Para no repetir palabras
Evitar ser redundante a la hora de estarse expresando
para demostrar uno de sus pensamiento que tubo en el pasado
Para hacer referencia a algo mencionado con anterioridad.
para sonar con un lenguaje más formal, no tan coloquial
para dar una conotación
Para referirse a un objeto de una manera menos coloquial
en estos casos se usa para describir de una manera despectiva a una cosa o persona
Porque se ha normalizado esa forma de usar l palabra
Falta de vocabulario
Para enfatizar el sentido de la oración
por que se refieren a objetos en mi caso asi pasa
"dicho" suena a texto traducido, aunque es correcto no es coloquial.
Pretender ser mas intelectual
Nada en específico. Solo se usa cuando hay que utilizarlo
Para que la oración se escuche más formal/mejor.
"Dicho" tiene como objetivo no repetir un sustantivo previo.
Para expresar mejor lo que intentan decir
para reafirmar lo que acabas de decir
Para referirse a lo que anteriormente se dijo y no repetir
Formalidad
Para hacer más énfasis sobre lo que están hablando
para parecer interesantes tal vez

yo creo que se utiliza mas por costumbre decirlo de esta manera
Creo que es para hacer enfasis en el mensaje (o remate del mismo) que se quiere expresar.
Para parecer más inteligente
Para referirse a lo que menciono anteriormente sin decir las cosas dos veces
Para no repetir el evento al que se están refiriendo. Es una palabra formal y no del habla cotidiana
Para especificar a lo que se están refiriendo y no dar lugar a malentendidos.
puede ser para, parecer un poco mas inteligente
Querer reforzar su creencia de que el elemento en cuestión es común para los demás o para no describir nuevamente al elemento.
Como una palabra despectiva
I think they only use it as if they used some other word, I do not think they have any purpose in using it
Para dejar en claro sobre lo que se habla, es algo que sin darnos cuenta se volvió parte del vocabulario de los usuarios de redes sociales así como utilizar "like" al describir lo que otra persona dijo en una anécdota
No
no pienso que sea a propósito, mas bien es una costumbre si alguien ocupa regularmente esa palabra
Es la manera mas propia de usar la palabra en una oración
Es una palabra muy poco utilizada pero puede ser por tener muy buena noción de nuestro lenguaje.
para referirse a ese objeto antes mencionado
Para sonar mas elocuente
No lo sé, no lo había pensado
para recalcar que ya se habló del tema (el sujeto de la oración)
Lo usan como muletilla
Sí, señalar o recalcar que hablan de una particular cosa.
normalmente para referirse a determinado objeto y hacer énfasis en el
Para mencionar algo previamente mencionado de una manera que parezca más

sofisticada.
Para referirse a algo mencionado anteriormente y no tener que volver a repetirlo
la gente suele decirlo y a la gente se le queda grabado en su mente.
Realmente no creo que muchos la usen más que los que quieren sonar inteligentes
A veces suena demasiado formal
creo que la gente lo usan para especificar el asunto o especificar que sucedio eso con ese objeto
Para ser más formal
Para farolear.
Creo puede ser para sentirse más sofisticados
son costumbres de cada persona o pais
Para hacer énfasis en el objeto o persona
creo que es por modismo
Probablemente para sonar con mayor seriedad
Pienso que es quieren presumir que saben hacer uso correcto del español
Lo hacen en "referente" a algún tema
para especificar que es ese cafe en especial y no cualquier otro
Para referirse de una mejor manera al objeto del que se esta hablando
Para ejemplificar una acción de una mejor manera.
para reafirmar informasion
para describir un lugar
Es una manera de evitar los vicios del lenguaje.
Supongo que sí, ya que así no se usa redundancia
no lo sé tal vez para sonar mas "inteligente"
Es una palabra que permite hacer énfasis sobre el objeto del que se está hablando, le da una especie de importancia
Una manera más formal de decir las cosas, nada más.
Para expresarse de forma más sucinta
I have no idea

Es una manera de referirse a lo que se escribió previamente sin volver a escribir la palabra en cuestión.
no realmente, es como cuando dicen literalmente y no significa nada, solo para hacer la oración más inteligente
Para no repetir el objeto, y que no suene raro la oracion
No creo que haya una razon, simplemente asi es el lenguaje
para referirse a algo
Para hacer énfasis
Para que la oración se escuche de una forma más formal y seria
gramatical rules
no tengo idea de porque se use de esa manera, el proposito es hacer enfasis en lo que se dijo anteriormente, aunque se vuelve redundante.
se utiliza para confirmar algún argumento
quizás
Para remarcar al mencionarlo
Para hacer énfasis
para expresar algo ocurrido en ese momento que no puede dejar de pasar
Sonar de una manera intelectual.
para hacer un personaje a un objeto inanimado en una historia o para referenciar algo dicho anteriormente
Para ser mas explicito en la idea
Tratar de simular un lenguaje educado / culto
para que la gente que esta leyendo entiendan a lo que se están refiriendo
para no sonar redundante o sonar intelectual
Es parte de la gramática del español
para dar enfasis
No lo se
No tener que repetir la misma palabra
No lo hacen
Tiene sentido en el contexto y es correcto gramaticamente

es como conector de las palabras
Para variar el texto en la oración
Para no repetir lo que ya dijeron
N/A
Supongo que el proposito es para resaltar el objeto directo.
No creo
no creo que tenga un proposito en particulas
Siendo en alguna red social, creo que la usan solo por querer demostrar que son muy cultos o algo así.
Para reforzar el sustantivo en una oración
Para que parezca más formal
para referirse a algo
Es una manera fácil de recalcar que te refieres a algo que ya mencionaste.
Es alusivo a la cosa, situacion o persona. Transmitir un mensaje sin ser redundante.
Para clarificar a que se refieren
Refererise a algo mencionado previamente
La gente no suele usarlo, porque su forma de expresarse es más informal.
Para sonar mejor, es lenguaje rebuscado
Para referirse a un o bjeto
Pienso que suena un poco mas dramatico
Creo que lo utilizan para enfatizar un cierto aspecto de su oración, siento que es un tanto formal, pero supongo que resalta la idea.
En un uso común de dicha palabra, de dicho concepto. No está mal utilizarla así, forma parte de nuestro lenguaje. Es un estilo de escritura, una forma de expresar una idea.
No creo que se uitilce mucho la palabra "dicho" en redes sociales, ya que es ligeramente formal
Hacer énfasis sobre lo que se está hablando
para sobre-especificar
para ser más concretos
Para no repetir toda una oración

Es una forma coloquial de comunicarse
Es una buena forma dar énfasis en el objeto del cual se habla en dicha oración, dando un contexto mas propio y formal a dicha oración. ¿Ven?
Pienso que es una forma de reafirmar el objeto del que habías hablando con anterioridad.
para resaltar al objeto, persona o acción que se refiere
Para enfatizar algo
probablemente para dar variedad a la escritura, para que no se vea monotono.
Es para referirse a una cosa o suceso de forma más clara.
Para darle un poco menos de repetición a las palabras y hacer énfasis en la situación.
Para no repetir la misma palabra.
Por que es útil, aunque no es tan común.
Para reiterar que se habla del mismo que se mencionó
Muletilla similar a "el cual/lo cual"
Describir el objeto de la oración
Impresionar tal vez, no lo sé.
Es una manera mas educada de hablar y varios quieren parecer así
Creo que se usa el término "dicho" para referirse al sujeto de la oración y así no repetir el nombre del sujeto en cuestión.
Utilizaran Dicho, por costumbre, en Mexico, es poco comun, pero tal vez en España me imagino que si lo es.
Hace referencia al objeto o situación de la que se está hablando
Referencia gramatical redundante, muy habitual, no podría ser considerado error, ni acierto
Tengo entendido que es una formación gramatical totalmente correcta para referirse a algo o a veces a alguien que ya se ha mencionado con anterioridad y que no se quiere volver a repetir los detalles para hablar de forma más rápida o fluida. Básicamente quiere decir "sobre el que ya se ha dicho o hablado".
Buen manejo del lenguaje
No
suena muy propio, para sonar elocuente

No lo sé
Para no repetir la palabra a la cual se refieren
lo usan como un sujeto, en lugar de el

APPENDIX L

Social meaning experiment:
Task 2 question 6, Spanish study

Complete la siguiente oración: **Yo pienso que la gente usa "dicho" cuando quieren parecer _____.** **(Su respuesta puede ser más de una palabra si es necesario)**
Inteligente
ortográficamente correctos
intelectual
Interesantes.
interesante
Mas imponentes
intelectuales
interesante
Más inteligentes y/o educadas
Interesantes
culta, inteligente
formal
especifica
formales
tener un buen léxico.
Diferentes, auténticos
inteligentes
inteligentes

demasiado propios.
nada en especifico es una usansa
Propios al expresarse.
snarky
Formal
Educada
sofisticada
Autoritario.
amigables
Fuertes, decididos
cultos
formales, educados.
graciosa y dominante, tajante también.
Cómunes
culta
graciosa
contundentes
Intelectuales
rebuscada
formales
Mas inteligentes o cultos
más interesantes de lo que son.
inteligentes
interesante
intelectuales
cultos
inteligentes
Inteligentes
Cultos

Ni idea, no me parece importante
elegante
Culta
Inteligentes
personas con varias formas de expresarse
llamativas
conciso
inteligentes
mas sofisticados
con un vocabulario más amplio
desinteresada
Sofisticada
Más interesantes
Razonable
personas que se refieren a cosas no propias
Pretenciosos
inteligente
Correcta
Más inteligentes
culta
Inteligente
Mas intelecutales
Interesante
Formal
Más específicos sobre una cosa o tema
sofisticados
interesantes
Un poco ingeniosos, cómicos u ocurrentes...

Inteligente
que tiene un mejor conocimiento del lenguaje
formal
interesantes
interesante
Confiables, conocedores.
Interesante al darle al "objeto" un valor mayor o para que destaque
a little more interesting
Pasivo agresivo respecto al sujeto de la historia
intelectuales
mas inteligente de lo que es
inteligente
Ostentosos
elocuente
elocuentes
Inteligentes
una persona de extenso vocabulario
inseguros
Cultos
de amplio vocabulario
elocuentes
Interesante o culta
interesantes
Inteligente o interesante tal vez culta
Educados
inteligente
Formales
Pedantes.

Refinando
practicas
Importantes
no lo se
Serios
intelectuales
Más intelectuales
mas intelectual
Intelectuales
Más claros.
redundante, porlo tanto inteligente
intelectuales
más intelectual
Más inteligente, o formal
culto
inteligentes
profesionales
Graciosos
geniales
Cultos
inteligentes
que tiene lexico
formales, cultos, propios
inteligentes
Intelectual
culta
intelectual
intelectuales
mas seguros en lo que dicen

Cool
mas enfaticos
Eruditos
interesante
Intelectuales
graciosos
intelectuales
Formales
educados
Intelectual o buen redactor
Inteligentes
intelectual
Simpaticas o burlonas
inteligentes
Una persona que no existe
formales
dar mas enfasis
Que tienen buena redacción
Inteligentes
novedosa
Redundantes
interesantes.
nada, simplemente una manera de hablar diferente
Inteligentes, cultos
Nada, creo que no es tratar de parecer algo...
Intelectuales
mas formales
cultos

Breve
Muy explicitos
pretenciosos
Inteligentes, intelectuales o formales.
Intelectuales o interesantes
Normal
interesantes
educada
graciosos o propios.
más formales
más cultos de lo que son
mas especifica de lo usual
Intelectuales
Mas inteligentes
Inteligentes
Un poco mas diestra en el uso del español.
Inteligentes
cultos
Interesantes
interesantes
Este o esta.
más inteligentes que el resto.
interesantes o de un amplio vocabulario.
Clara
interesantes
intelectuales
Cultos
Inteligentes
Educada

Sarcástico, coloquial, interesante.
español
cultas
intelectuales
más inteligentes y precisos.
Culta
Intelectual
letrados, que hablan bien.
Interesantes
Menos repetitivos
cuando quieren parecer serios pero de una manera un poco sarcástica

APPENDIX M

Social meaning experiment:
Task 2 question 7, Spanish study

¿Cómo cree que “dicho” sea distinto en su significado entre otras palabras tales como “el” o “la” o “ese” o “eso”? Por ejemplo, “Compré un café en Starbucks, pero enseguida derramé ese café en mi suéter nuevo.”
Es una manera más elegante de decirlo
suena menos formal
Le da mas formalidad a la oracion
Es lo mismo.
"dicho" suena más formal
Dicho se referiria hablando ya de un cafe antes mencionado
Al leerlo da una sensación de formalidad
compre y cafe y se cayo en mi sueter
No tiene diferencia
Dicho suena más interesante, da más importancia al artículo del que se habla
parece ser el mismo significado
cambia el significado dependiendo del contexto.
mas formal
es decirlo impersonalmente, como si no estuviera ahi
ES UN POCO MAS PRETENCIOSO
Es diferente en el aspecto de que remarca mejor de lo que se está hablando
Suena más formal.
no es distinto es otra manera de decirlo

"lo": lo derramé.
si
Tiene un uso menos común que los otros pronombres mencionados ("el", "la", etc...) y, por lo mismo, supone un vínculo formal con el sustantivo al que alude.
No sé
Dicho me parece una palabra más profesional
No creo que realmente haya una gran diferencia al tratar de entender la idea.
suena mejor decir "dicho"
Igual - Mismo Significado - Quizás Por el País Donde Se Use. Dicho Suena Raro. El Para Mi Suena Correcto.
suena como al referirse a algo ajeno de manera cercana
Es una palabra o frase más sensible
no
No creo que sea distinto, solo que suena formal.
Tomando el ejemplo probablemente quieran verse con un vocabulario un tanto más refinado según ellos
Es lo mismo.
igual
se refiere a otra cosa, o le da diferente contexto a la oración
intenta hacer un enfoque mas especifico sobre algo
Dicho suena más propio mientras que ese es un adjetivo que denota un lugar
suena mas corriente
Le da otra connotación y al ser más formal, llama la atención del receptor.
Que dicho se utiliza cuando ya has declarado un sujeto.
Dicho es usado como si se dijera "el mencionado anteriormente".
Nada
una forma similar
es más especifico
Solo es otra manera mas compleja de decirlo
la gente quiere parecer más inteligente al utilizar esta palabra que no es tan común.

Es igual pero más rebuscado
Solo suena mejor
En ninguno
Más propiamente, o más elegantemente
Puede ser mas útil para oraciones mas especificas y complejas.
Suena mas formal
No hay mucha diferencia en realidad, solo es cuestión de redacción.
puede ser lo mismo depende del contexto
Hace un mayor énfasis y es más claro que usar un pronombre vago como "él"
en que no señalas algún objeto mediante las palabras, más bien, retomas algo que ya mencionaste
compré un café en Starbucks, pero enseguida derramé ese dicho café en mi suéter nuevo
No creo que sea especialmente diferente
no tiene tanto impacto en la oracion usar otras palabras ya que dicho suena mas despectivo
No creo que semánticamente sea distinto, solo suena más propio
Para ser más específico
No
no es distinto
es un lenguaje mas sofisiticado contra uno mas relajado
No es diferente, tienen el mismo significado
No lo sé
Se escucha más formal
Su uso es un poco más culto y, en ocasiones, un poco presuntuoso.
Es más inexacto y se puede utilizar para más objetos sin determinar el sexo.
Suena a que la persona es mas culta
"dicho"se refiere a lo que se está hablando y no a lo que se observa como "ese café"
Dicho" es para situaciones más formales
No es distinto

su significado no es muy diferente pero no es muy común emplear esa palabra
yo creo que es el mismo significado pero dando a entender que se hablo de el anteriormente
Al usar “dicho”, se hace a ña expresión mucho más personal que si usaramos otros artículos o adjetivos.
Dicho es un término más formal mientras que la, el o ese suenan más del día a día
Solamente es una forma de decirlo
Dicho es mas formal
Creo que trata de separar más el objeto de la persona que lo enunciado, para no ser relacionado con ello.
por que dicho es para definir una frase o algo que alguien dijo alguna vez
Refuerza la intención de señalar el elemento principal, en este caso, el café.
Siento que se usa como una palabra de cirta forma despectiva
that can be used in other types of sentences giving different meaning
Le da un mayor énfasis al sujeto
estas refiriendote a algo que mencionas en la misma frase, mientras que las otras palabras son mas generales
pertenencia o el momento de explicar
N/A
Le da más clase a la oración usar "dicho"
es el mismo significado, solo que dicho es más elegante
Para dar enfasis
No cambia
mismo significado
ese
Dicho; es más formal y da a entender que la oración esta centrada en un sujeto específico sin sonar redundante.
se le intenta dar mayor importancia a la frase
"dicho" nos remite a textos académicos, y es poco común encontrarnos con esa palabra en nuestro vocabulario del diario, a diferencia de otras palabras mucho más comunes.

Como para resumir un poco lo anterior comentado y se de a entender a lo que se refiere
que menciona justo al objeto antes mencionado.
Realmente no cambia mucho y creo que incluso sería mejor noche usarlo
Ese suena más casual, apropiado para una red social
se usa de manera igual
Hace que la oración sea más formal
Es igual, pero sería mejor decir "y enseguida lo derramé en mi suéter nuevo", pues es redundante.
Para mi es lo mismo solo suena mas propio
es una costumbre utilizar la palabra aunque no este bien utilizada
El dicho a mi parecer lo usan para hacer enfasis
no es distinto
Gramaticalmente "dicho" es más agradable
las segundas palabras son más informales y se usan en el día a día, a diferencia de "dicho"
Ese
pues creo que sale sobrando la palabra ya sea dicho o ese
Solo como palabra puede usarse de las misma manera como las demas para referirse a algo o alguien
No suena tan apropiado, ni tan elegante.
hace mas pomposa la oracion
es mas formal
No es diferente pero la palabra dicho es un poco más formal que simplemente utilizar un articulo para referirse al cafe
Suena mas formal
me suena a lo mismo pero de una forma mas elegante
Permite hacer más énfasis en las oraciones
Dicho es más específico.
Obliga al lector a reflexionar sobre el inicio del relato

mas original
Es más específico
llama más la atención
no en el significado pero si en como suena
Es mas formal, ademas hace que suene menos redundante la oracion
para usar dicho normalmente tienes que haber mencionado antes a lo que te refieres
Suena más formal
es más preciso que decir las otras palabras
I do not know
tiene menos seriedad y suena mas casual.
cambia el significado porque es para señalar
Muy distinto
no es distinto en realidad
Enfatiza más las palabras dentro de la oración.
si puede ser distinto depende de la oracion que se desea emplear o dar a entender
Es muy parecido
el dicho es mas como un modismo, los demas solo son los hechos
No mucho, podría considerarse un sinonimo
No me parece que de algún cambio, pero sí que de más "elegancia" a la frase
es mas generalizado el termino
porque "dicho" se puede usar el lugar de todas esas palabras
Suena más pomposo
dicha palabra das mas enfasis
Le otro tipo de significado a la oracion
no cambia mucho el significado
No lo es, es un participio flotante
no es distinto, solo le da un tono diferente a la oracion
siento que lo usan para dar mas enfasis o un toque mas intelectual
No muy distinto

Con “dicho” no hay que repetir el sustantivo del que se habla
N/A
Creo que solo es una forma de variarlo.
no
no tiene ningun significado diferente
Pues es igual
Sinonímo...
Una forma más elegante de decirlo
yo creo que significa lo mismo pero al usar dicho parece mas formal
Yo siento que "dicho" es mucho más específico, especialmente en oraciones o conversaciones más largas donde se puede confundir de qué estás hablando.
El uso de ese, muestra el momento el objeto - vaso de café- visible.
Es un uso mucho mas coloquial
Suena más natural.
Se ve más elocuente
Suena más “elegante”
No ed
Suena mas simplon
La palabra dicho me comunica un sentido de formalidad que "el" o "ese" no me comunican. Parece un poco más inusual.
Es una expresión o palabra más propia, en cierto sentido. También sirve para enfocar la atención en los objetos de otra forma, he visto la figura utilizada en chistes muchas veces.
Suena más casual no utilizarlo
Que al escribir "dicho" se da otra connotación graciosa al suceso mientras que los demás artículos no
menos caracteres, mayor entonación
Es más coloquial
Los españoles lo usan mas, o personas que quieren ser correctas
Es un sinónimo

Dicho hace la función de pronombre, haciendo redundante la necesidad de repetir el nombre completo del objeto al cual hace mención
No creo que haya diferencia en ese contexto.
porque se refiere a ESE café
No parece distinto en el uso
tiene el mismo significado
Es más formal
Es un lenguaje más formal.
Dicho suena más formal y da la impresión de que la persona tiene un vocabulario más amplio.
Nunca lo había pensado, pero creo que es completamente intercambiable por las opciones presentadas
Pienso que dicho no es diferente pero representa una falta de conocimiento de artículos
Es más correcto con "ese" que con "dicho"
N/A
La verdad no estoy seguro, pero no me agrada la palabra.
Dicho es un termino mas rebuscado.
Pues, según el estado de ánimo, puede indicar lo que siente la persona.
No se
Es más especifico que los artículos "el" "la".
Solo es para referirnos al sujeto previamente escrito
Yo he notado que cuando se habla o se trata de varias cosas, objetos o elementos en una conversación, y no se habla solamente de una cosa u objeto, es cuando se nota la mayor y verdadera utilidad de esa expresión, pues de algún modo al usar dicha expresión o ésta expresión se busca que se haga referencia a un objeto del que ya se ha hablado con anterioridad. Esa expresión es de uso muy frecuente en el idioma español y al parecer también cumple la función de no repetir demasiado la palabra ése, éste o aquel, pues para evitar también la repitición constante de las mismas palabras o expresiones.
Cita al sujeto mencionado antes
No se

Siento que está mejor escrito, le da enfasis en el sujeto
Es lo mismo
Depende el contexto en que se este usando
No estoy seguro, pero creo que dicho no se puede usar para intercambiar un pronombre