Informativity and Linearization in Reference Production

Muqing Li

an der Philosophischen Fakultät

der Universität des Saarlandes

Dissertation zur Erlangung des akademischen Grades eines

Doktors der Philosophie

2025

Disputationsdaten

Betreuer:	Prof. Dr. Matthew W. Crocker
Erstgutachter:	Prof. Dr. Matthew W. Crocker
Zweitgutachter:	Prof. Dr. Ingo Reich
Vorsitz der Prüfungskommission:	Prof. Dr. Alexander Koller
Datum der Disputation:	01. Oktober 2025
Promotionskommission:	Prof. Dr. Matthew W. Crocker
	Prof. Dr. Ingo Reich
	Prof. Dr. Alexander Koller
	Prof. Dr. Felicitas Kleber
	Dr. Francesca Delogu

Acknowledgements

I would like to express my gratitude to my supervisor, Prof. Dr. Matthew Crocker, for recruiting me into the C3 project Rational Encoding and Decoding of Referring Expressions of SFB1102 Information Density and Linguistic Encoding. I am especially grateful for his support and allowing for me to pursuing the challenging and underexplored field of language production. I also thank him for his comments on this thesis.

I am grateful to my collaborators, Dr. Noortje Venhuizen, Dr. Heiner Drenhaus, and Dr. Torsten Kai Jachmann. Noortje and I jointly developed the innovative maze-based sentence completion task to record participants' referential choices, which was a pivotal moment enabling this project to continue while we were lost in repeated observations of the overwhelming pre-nominal modifications across multiple experiments and spoken tasks. Torsten proposed to use identical visual displays across experimental conditions, improving control over visual complexity and salience, which was critical but less concerned before. Torsten and Heiner also offered substantial support in statistical analysis and experimental design. Together with Prof. Crocker, they all provided insightful feedback that helped shape the relevant publications of this project. I also thank Sahra Slet for her assistance with Experiment 5.

I am thankful to my mentor, Prof. Dr. Bistra Andreeva, for her guidance and support throughout the progression of my PhD. I also extend my thanks to Prof. Dr. Elsi Kaiser, Prof. Dr. Jennifer Arnold, and Dr. Elli Tourtouri, for their in-depth discussions and advice on this project, during their academic visits to Saarbrücken and at conferences.

I would like to acknowledge the Committee Members for their suggestions and feedback in improving this thesis.

Special thanks to the SFB coordination team, especially Marie-Ann Kühne and Claudia Verburg, for their administrative support, some of which was tailored to accommodate my unique situation as a non-German resident.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project-ID 232722074—SFB/CRC 1102.

The thesis was written using the *Oxforddown* Template (Lyngs, 2019) in RMarkdown. All statistical outputs and figures were generated directly through R code while compiling the thesis. I am grateful to the developers of these tools for making this writing process manageable.

On my personal level, it was a difficult journey. For many reasons, I was not socially active during the seven years of my PhD, but this does not mean that my life here was in total isolation. I am deeply grateful for the meaningful friendships I formed with people who supported me throughout this journey. I

thank the Saarmegen team (a blend of *Saar*brücken and Nij*megen*), Dr. Marjolein van Os and Dr. Christoph Aurnhammer, with whom I shared both the Master's experience in Nijmegen and the PhD experience in Saarbrücken. I am also grateful to know Dr. Wei Xue, whose academic path intersected with mine in three cities, from Tianjin to Nijmegen to Saarbrücken, which is a rare and special overlap. I also thank my dear SFB colleagues, including but not limited to Dr. Omnia Ibrahim, Miriam Schulz and Miaoran Zhang, for their support and company in daily work life.

My history with Saarbrücken was longer than this PhD: in 2015, I spent a bachelor's exchange year here. I am thankful to still be connected with friends from that time, as well as to many kind and supportive friends throughout the PhD. I especially appreciate the help I received during 2024, when I was temporarily immobile due to an ankle injury. I also want to acknowledge the continued friendship of those from my hometown, who stayed in touch daily despite the distance and time zones.

This PhD journey coincided with an exceptionally challenging period in my life. In addition to the academic demands, I navigated complex personal circumstances and global disruptions, including the Covid pandemic. These years required sustained strength and adaptability, and I am deeply grateful to both the professional support system and the individuals who helped me navigate these obstacles, as overcoming even one of them would already have been a great accomplishment on its own.

In the end, I give credit to my persistence and resilience for carrying me through this journey.

Muqing Li

李沐晴

Abstract

In visually-situated referential communication tasks, speakers must select relevant visual properties and determine their linear order within a syntactic structure in order to encode a message that enables the listener to successfully identify the intended referent. While previous studies have primarily focused on the influence of informativity on property selection, especially overspecification, little is known about how informativity affects the linearization of property order, particularly when syntactic variation is involved.

This thesis investigates whether and how the informativity of property words, as determined by visual-situated contexts and quantified via Referential Entropy Reduction, influences syntactic linearization. Five referential communication experiments investigate whether informativity modulates speakers' syntactic choice between pre-nominal and post-nominal modifications in German, when describing referents in visual scenes depicting animals performing actions. Additionally, the project explores the role of communication engagement by reinforcing perspective-taking and comparing web-based and face-to-face interaction settings.

The results reveal two groups of speakers: *Group Consistent*, who are insensitive to informativity and adhere to a fixed syntactic structure, in line with a speaker-oriented, heuristic production approach; and *Group Varied*, who vary the use of syntactic structures to adjust property orders based on informativity, favoring an informative-first linearization strategy that facilitates target identification for listeners. The proportion of *Group Varied* speakers increases with communication engagement, particularly in the most engaging face-to-face interactions and when perspective-taking is reinforced.

This thesis advances our understanding of referential production and communication efficiency. Examining informativity, a speaker-external, listener-oriented factor, provides a clearer distinction between the speaker-oriented and listener-oriented views of reference production, as reflected in the different linearization strategies adopted by the two speaker groups. The distribution of these two groups is mediated by perspective-taking and communication engagement. The informative-first linearization preference offers novel evidence for the role of communication efficiency and audience design in shaping syntactic choices during the early grammatical encoding phase of language production.

Keywords: informativity; reference production; linearization; syntax; communication efficiency

Contents

List of Figures

List of Tables

List of Abbreviations

AIC . . . . . Akaike Information Criterion
BIC . . . . . Bayesian Information Criterion
DO . . . . . Double Object
FPL . . . . . First Property Latency
HHC . . . . . Human-to-Human Communication
HCC . . . . . Human-to-Computer Communication
MS . . . . . Minimal-Specification
OS . . . . . Over-Specification
PO . . . . . Prepositional Object
RER . . . . . Referential Entropy Reduction
RSA . . . . . Rational Speech Act
UID . . . . . Uniform Information Density

1 Introduction

Contents

In daily conversations, referring to objects is a fundamental function of language. Imagine you're at a bakery and need to specify which cookie you'd like to buy from a selection laid before you, as seen in Figure 1.1. You might say, "I would like to buy *the flower-shaped cookie*." The waiter, after knowing which cookie you mean, would respond, "Sure", and hand the cookie to you.

This type of communication is known as **referential communication**, where two interlocutors, a **speaker** (you) and a **listener** (the waiter), exchange information about a visual scene like Figure 1.1, containing multiple referents sharing and differentiating certain visual **properties** (e.g., color, shape, flavors and type of the cookies). The speaker often focuses on describing one referent, or the **target**, while other referents serve as the **context**. The utterance about the target produced by the speaker is a **referential expression**.

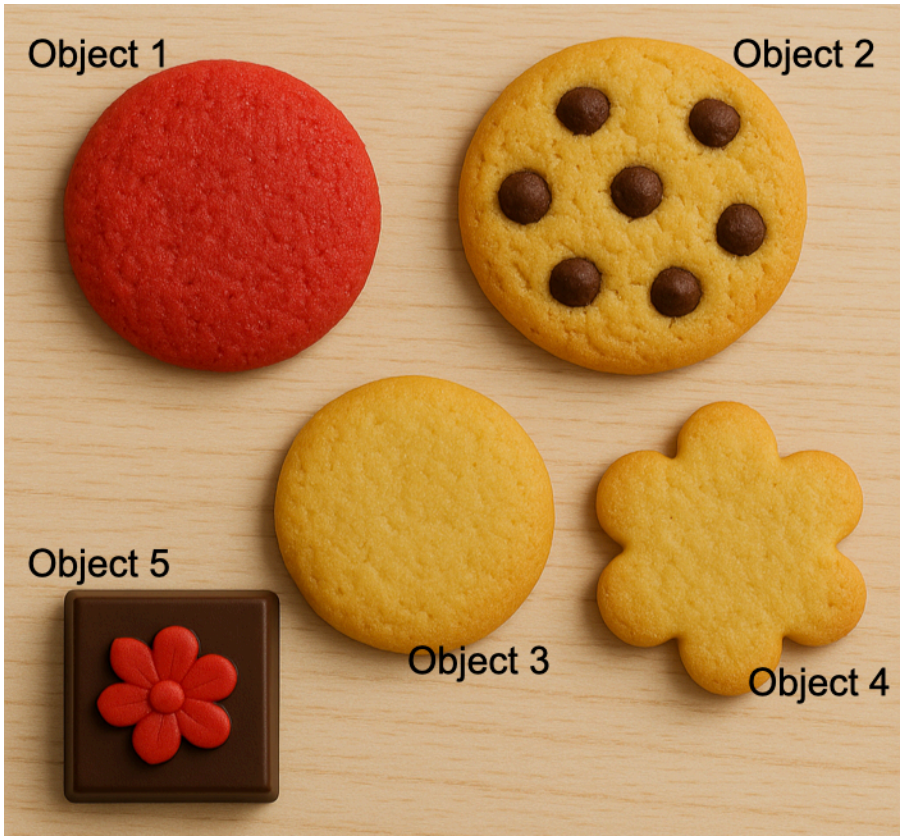


Figure 1.1: An example of a visual display for referential communication. The picture is AI-generated using OpenAI's DALL.E, via ChatGPT (OpenAI, 2025).

In referential communication, speakers must make several decisions when formulating referential expressions. First, speakers need to choose which properties of the target referent they will mention, a decision regarding **property selection** (e.g., Belke & Meyer, 2002; Dale & Reiter, 1995; Gatt et al., 2013, 2017; Koolen et al., 2013; Pechmann, 1989). Consider Object 1 in Figure 1.1, the red round cookie. You might describe it as "the red cookie", "the round cookie", or more explicitly as "the red round cookie". The expression "the *round* cookie", which selects only the *shape* property but not *color*, however, may lead to ambiguity, since multiple cookies share the same round shape in this visual context.

In addition to deciding which properties to mention, speakers also need to decide how to arrange the selected properties within the expression, which concerns **property ordering** (Fukumura, 2018) or **linearization** of referential production. For example, to describe Object 4 (the flower-shaped cookie), you could say "the flower-shaped cookie" or "the cookie with a flower shape". Both descriptions convey essentially the same set of properties, yet differ in the order these properties appear. In addition, these ordering differences may also entail syntactic differences, in the current case, between pre-nominal and post-nominal structures.

What factors would influence speakers' decisions regarding property selection and ordering? From the **speaker's perspective**, expressions that are easier and more habitual to produce might be preferred. English speakers, for example, would

typically use the pre-nominal structure (e.g., "the red cookie"; "the flower-shaped cookie"), rather than the post-nominal structure ("the cookie that's red"). The speaker may also focus on properties that stand out visually. For instance, one might quickly notice the salient color "red" of Object 1, making it more likely to be mentioned.

From the **listener's perspective**, waiters from the bakery would probably expect referential expressions that are concise and helpful for them to accurately distinguish the intended cookie from the rest. To facilitate this identification process for the listener, the speaker ideally should mention properties that can best distinguish the target referent from other objects in the visual context. For instance, saying "the flower-shaped cookie" clearly identifies Object 4 because the *flower-shape* uniquely distinguishes it from the others. By contrast, mentioning other less distinguishing properties, such as color in "the *yellow* flower-shaped cookie", is less helpful, as multiple cookies share the same color *yellow*. This may leave the waiter, as the listener, uncertain about which exact cookie the speaker refers to when just hearing the color "yellow".

Additionally, from the listener's perspective, the most distinguishing property should also ideally appear early in the referential expression, allowing the listener to identify the target efficiently, potentially even before the speaker finishes speaking. Thus, the pre-nominal form "the flower-shaped cookie" might be more efficient than the post-nominal alternative "the cookie with a flower shape" for Object 4 in Figure 1.1, since for the post-nominal expression, the waiter would have to wait until the end of the expression to receive the crucial distinguishing information "flower". By contrast, when referring to Object 5, it may be more efficient to begin with the noun "chocolate," forming a post-nominal expression such as "the chocolate topped with a (red) flower." This is because "chocolate" uniquely identifies Object 5 in the display, as there is only one chocolate, whereas the property "red" (also describing Object 1) or "flower" (also describing Object 4) are shared with other referents. In this case, the listener can already identify the target upon hearing "chocolate", making the post-nominal structure more efficient than pre-nominal alternatives like "the red chocolate" or "the flower-topped chocolate."

The extent to which a property distinguishes the target from other objects relates directly to the concept of **informativity**, which appears to be a crucial factor shaping speakers' referential expressions in the example above, particularly when speakers actively consider the listener's perspective in deciding property selection and property ordering (e.g., Fukumura, 2018; Rubio-Fernández, 2016; Tourtouri et al., 2019).

The current thesis investigates precisely this influence of informativity on speakers' property ordering decisions in referential expressions, i.e., the linearization of multiple properties. While property selection has received extensive attention in previous research, revealing that speakers often include more properties than necessary (i.e., overspecification: e.g., using "the *yellow* flower-shaped cookie," even though shape alone is sufficient. Koolen et al., 2011; Pechmann, 1989; Rubio-Fernández, 2016; Sedivy, 2003; Tarenskeen et al., 2015; Tourtouri et al., 2019, among other studies), much remains unknown about property ordering, especially when different orders entail syntactic differences (e.g., pre-nominal vs. post-nominal

modification). Existing research on the role of informativity in property ordering is limited (except for Fukumura, 2018; Haywood et al., 2003), leaving open questions about whether and how informativity modulates speakers' choices of property orders.

This introductory chapter begins by situating the referential communication task within the broader framework of an information transmission system, highlighting the importance of communication efficiency (Section 1.1). A central aspect of communication efficiency is **informativity**, which is introduced as the main focus of the thesis, and it will be defined from both qualitative and quantitative perspectives in visually-situated contexts (Section 1.2). The chapter then turns to the question of how referential expressions are planned and produced by speakers. Section 1.3 provides an overview of the linearization process in referential production, placing it within the broader language production process shaped by both speaker-internal mechanisms and speaker-external, communication-driven factors such as informativity. Finally, this chapter provides an overview of the thesis structure and the research questions addressed in the project.

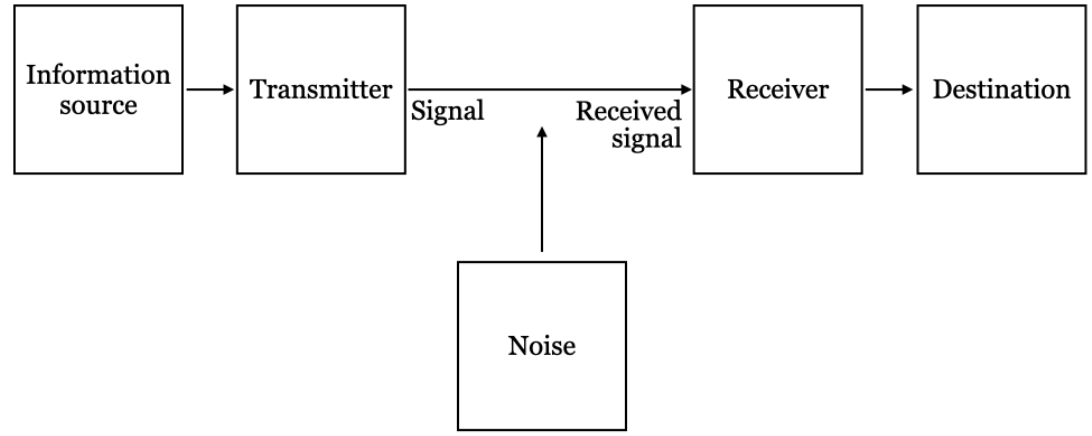
1.1 The Referential Communication Task as an Instance of Information Transmission

The referential communication process, illustrated earlier through the bakery example, has been formalized as an experimental paradigm known as **the referential communication task**, which has been widely used to study human communication processes (originally introduced by Krauss & Glucksberg, 1969, 1977; Krauss & Weinheimer, 1964, 1966). As shown in Panel B in Figure 1.2, the task typically involves two interlocutors — a **speaker** and a **listener** — communicating about **visual displays**[1] that contain multiple referents varying along several **properties** (e.g., color, shape and type). The speaker's task is to describe the target referent in a way that the listener can accurately identify it, often by clicking on the corresponding referent in the display.

Referential communication is a fundamentally collaborative process (Clark & Wilkes-Gibbs, 1986). Both the speaker and the listener are expected to act cooperatively (Grice, 1975). For the speaker, this **cooperative** process entails producing referential expressions that are accurate, informative, relevant, and clear — aligning with the four conversational maxims of Grice's Cooperative Principle: Quality, Quantity, Relation, and Manner (discussed further in Section 1.2.1). In the earlier cookie example, a cooperative speaker might use "the flower-shaped cookie" to provide accurate information that is sufficiently specific to identify the intended referent without excessive details. Conversely, non-cooperative speakers

[1] In the experimental task of referential communication, the locations of the referents are often different between the two interlocutors, preventing location-based descriptions (e.g., "the one on the right") and instead encouraging speakers to refer to object properties (e.g., "the flower-shaped cookie").

A. Information-theoretic model of communication from Shannon (1984).



B. Application to the Referential Communication Task

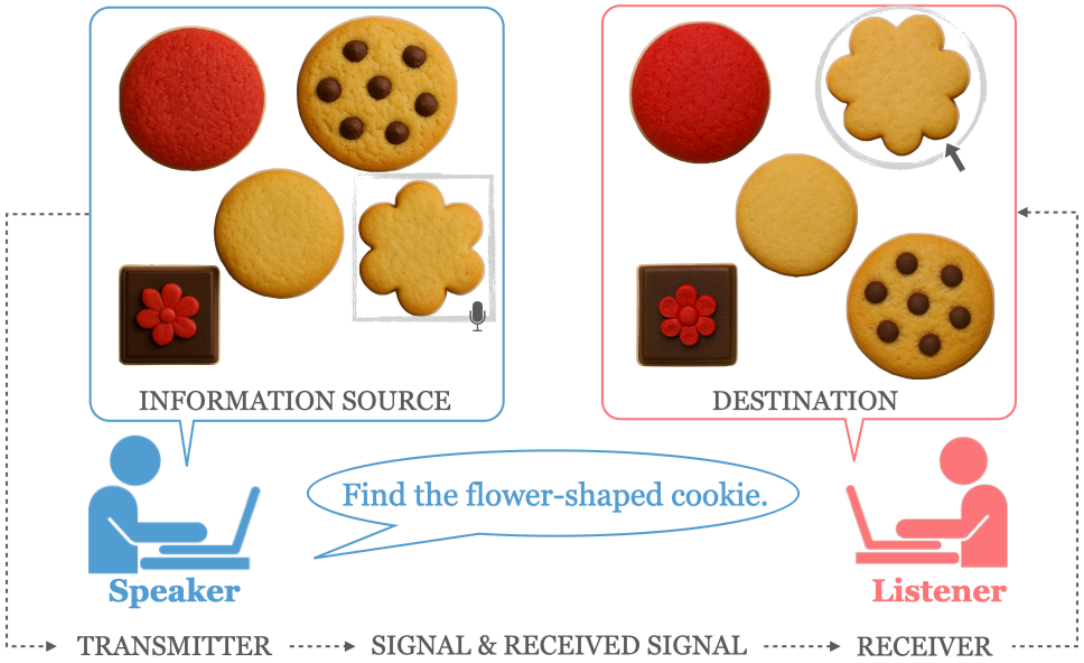


Figure 1.2: The referential communication task is an instance of information transmission in the framework of Information Theory (Shannon, 1948). Panel A shows Shannon's original model of an information transmission system for communication, recreated by the author. Panel B illustrates how this framework is applied in the referential communication task: the speaker, functioning as the transmitter, produces a referential expression of the target referent (marked by a grey square), such as "the flower-shaped cookie." The linguistic signal is transmitted to the listener, who, as the receiver, identifies the target by clicking on the selected referent (marked by a grey circle).

produce expressions that include inaccurate, unnecessary, irrelevant, or ambiguous information.[2]

In addition to being cooperative, referential communication should also be **efficient**. Although Grice's Cooperative Principle (Grice, 1975) touches on efficiency through the Maxim of Quantity (see Section 1.2.1), the notion of communication efficiency is more comprehensively addressed within the framework of Information Theory, as developed by Shannon (1948). Shannon (1948) conceptualized communication as a process of information transmission, adopting a rational, quantitative, and probability-based approach. In this framework, illustrated in Panel A of Figure 1.2, a communication system comprises a transmitter, a receiver, and a channel through which signals are transmitted. Inspired by telecommunication systems, Shannon's model highlights the rational dimension of communication, particularly in terms of efficiency, by focusing on how signals are encoded and decoded with respect to their probabilistic features.[3] An efficient communication system is one that achieves successful transmission of information with minimal average effort (Gibson et al., 2019), or, as summarized by Levshina (2022), by minimizing the "cost-to-benefit ratio". A classic linguistic example of this principle is the preference for shorter morphemes over longer alternatives, when both convey equivalent meanings in predictable contexts (e.g., chimp vs. chimpanzee, Mahowald et al., 2013; see also A. F. Frank & Jaeger, 2008).

Efficient speakers are expected to be **informative** when formulating referential expressions. According to the Maxim of Quantity (Grice, 1975), this means providing just enough information, neither too little nor too much. For example, in the earlier scenario, describing the target as "the flower-shaped cookie" offers just the right amount of details to distinguish it from other cookies in the display. This qualitative notion of informativity can also be quantified using measures derived from Shannon (1948)'s Information Theory, estimating the distribution of informativity values across linguistic units. The current thesis adopts this quantitative perspective in the context of the referential communication task, operationalizing informativity through **Referential Entropy Reduction (RER)**, a measure of how distinguishable a given property is in identifying the target referent (see Section 1.2.1, see also Tourtouri et al., 2019).

Comparing Shannon's communication model with the referential communication task reveals an analogy between the two (see also Sperber & Wilson, 1995). As depicted in Panel B of Figure 1.2, the speaker functions as the transmitter, encoding acoustic or textual signals based on the visual display and knowledge of

[2]These "general interaction principles" (Greenall, 2006) from Grice (1975) are better understood as a guideline for basic human communication rationality, rather than strict rules that speakers must always follow. In everyday conversations, these principles are, in fact, often violated — a phenomenon known as flouting (see, e.g., Greenall, 2006; Thomas, 1995 for further discussion).

[3]While highly influential, the model was not intended to capture all aspects of human communication. Other important dimensions, such as social intention, emotion (e.g., Kempe et al., 2013; Out et al., 2020) and non-verbal interaction (e.g., gesture, Robert M. et al., 1995), are largely outside its scope (see pragmatically oriented theories for these aspects in e.g., Levinson, 1983, 2024). These dimensions lie beyond the scope of Shannon's framework and are also outside the focus of the current project.

the target as the information source. These signals are then transmitted to the listener, who serves as the receiver responsible for identifying the intended referent (the destination). The communication channel may be the airwaves in face-to-face dialogue or a digital medium in virtual settings. While Shannon's model also includes a *noise* component, representing potential interference in signal transmission. Such disruptions are not directly addressed in the referential communication task in the current project, which assumes that the listener receives the intended linguistic signal without distortion.

This thesis conceptualizes the referential communication task as an instance of Shannon's information transmission system. It investigates speakers' referential encoding strategies and examines whether and how these strategies are modulated by informativity in the pursuit of communication efficiency. In particular, it focuses on an under-explored aspect of referential production, that is, the **linearization** of multiple properties within referential expressions, building on and extending existing research that has predominantly concentrated on property selection, especially on overspecification (see a review in Chapter 2).

1.2 Defining Informativity in Visually-Situated Contexts

As illustrated in the cookie example above, an informative expression, such as "the flower-shaped cookie" for referring to Object 4 in Figure 1.1, enables the listener (e.g., the waiter) to efficiently distinguish the target referent from other objects in the visual scene. This illustrates the notion of informativity adopted in the current project. Specifically, the term informativity refers to the extent to which an individual linguistic unit (here, a property word) contributes to distinguishing the target from other referents in a given visual context. This definition is consistent with how informativity is characterized both qualitatively in pragmatics (Grice, 1975) and quantitatively in Information Theory (Shannon, 1948), which will be introduced separately in the sections below.

1.2.1 Qualitative Definition of Informativity

Qualitatively, the notion of *informativity* (also referred to as *informativeness* in e.g., Davies & Arnold, 2019; Degen et al., 2020; M. C. Frank & Goodman, 2012) captures the amount of information a linguistic unit conveys in a given context. In this thesis, a linguistic unit refers to a property word. A closely related concept is *discriminatory power* or *discriminability*, which describes the extent to which a property word can distinguish the target referent from other objects in the visual scene (e.g., Fukumura, 2018; Gatt et al., 2013; Rubio-Fernández, 2021).

An expression is considered more informative, i.e., with a higher discriminatory power, when it narrows down the potential referents to a smaller subset of candidates in the context. For example, in Figure 1.1, if Object 4 (the yellow flower-shaped cookie) is the target, the shape property "flower" is more informative than the color property "yellow": "flower" distinguishes the target from all cookies

except Object 5 (the chocolate topped with a red flower), whereas "yellow" applies to three cookies in the scene. Similarly, if Object 1 (the red round cookie) is the target, the property "red" is more informative than "round", as "red" applies only to Object 1 and Object 5, while "round" refers to three different cookies. In both cases, property words with the greater discriminatory power, such as "red" and "flower", are more informative than less distinguishable properties of "round" and "yellow".

This definition of informativity highlights the interaction between linguistic expressions and the context in which they are used, as the informativity of a given expression may vary depending on the *contexts* (e.g., Davies & Arnold, 2019; Rubio-Fernández, 2021). Contexts can be categorized into *visually-situated contexts*, for example, referring to "the flower-shaped cookie" in Figure 1.1, or *discourse-based contexts*, such as mentioning "cookies" in a written article (see Davies & Arnold, 2019 for an overview of informativity in discourse-based contexts). The current thesis focuses on visually-situated contexts, specifically visual displays containing multiple referents as used in the referential communication tasks (e.g., Figure 1.1). In this setup, the informativity of referential expressions is related to the specific visual display provided to the speaker: the same linguistic expression (e.g., a word describing a visual property) may carry different amounts of information depending on the extent to which it distinguishes the target from the surrounding referents.

The qualitative concept of informativity has its roots in Grice's Cooperative Principle in pragmatic theories (Grice, 1975), specifically introduced in the Maxim of Quantity. This maxim defines the appropriate amount of information cooperative speakers should provide in conversation: speakers are expected to include as much information as is necessary, but not more than is needed. In the context of referential communication tasks, this Gricean principle can be further illustrated by three types of referential specifications: underspecification, minimal specification (MS) and overspecification (OS). These three types of specifications are defined based on whether properties mentioned are sufficient, insufficient, or more than required for uniquely identifying the target referent within a given visual context:

1. **Underspecification** occurs when the properties mentioned are not sufficient to uniquely identify the target. For example, in Figure 1.1, if Object 1 is the target referent, "the round cookie" would be an underspecification, as it could refer to Object 1 as well as other round-shaped cookies in the display. Such expressions fail to provide enough information for successful target identification. Previous studies have consistently shown that speakers are well capable of avoiding underspecifications that risk communication success (e.g., Brennan & Clark, 1996; Davies & Katsos, 2013; Davies & Kreysa, 2017; Engelhardt et al., 2006; Olson, 1970).

2. **Minimal specification (MS)** occurs when the properties mentioned are just sufficient — neither more nor less — to uniquely identify the target referent. For instance, "the flower-shaped cookie" serves as an MS for Object 4 in Figure 1.1, as it successfully distinguishes the intended object from all others in the display without including unnecessary information.

3. **Overspecification (OS)** occurs when more properties than necessary are included in the referential expression. For example, "the *yellow* flower-shaped cookie" for Object 4 is an OS, as the inclusion of the color "yellow" is not strictly necessary to identify the target. In such cases, the inclusion of additional properties is not in line with Grice's Maxim of Quantity, as it provides more information than necessary to identify the target.

Although MS may appear to be the optimal form of a referential expression based on Grice's Maxim of Quantity, a substantial body of psycholinguistic research has shown that OS occurs frequently in referential production (e.g., Koolen et al., 2011; Pechmann, 1989; Rubio-Fernández, 2016; Sedivy, 2003; Tarenskeen et al., 2015; Tourtouri et al., 2019, among other studies). When and why speakers produce OS frequently will be further discussed in Chapter 2.1, where informativity plays an essential role.

1.2.2 Quantitative Definition of Informativity

Quantitatively, informativity reflects the extent to which an expression reduces the uncertainty about the intended referent (e.g., M. C. Frank & Goodman, 2012; Jaeger, 2010). This aligns with the information-theoretic perspective of language (Shannon, 1948), in which the information content of each linguistic unit is determined by its probabilistic features. Two major metrics are commonly used to quantify informativity: Surprisal and Entropy Reduction.

Surprisal is defined as the negative logarithm of the probability of a linguistic unit given its context:

$$Surprisal(word) = -log_2 P(word|context)$$

Surprisal is a measure of informativity determined by the probability that an upcoming linguistic unit (here, a word) will occur given the preceding context. The higher the probability of a word, the lower its surprisal, the less informative the word is in the context, as it conveys information that has been largely expected by the previous context. Surprisal has been extensively studied in discourse-based contexts (e.g., Gibson et al., 2019; Jaeger, 2010; Jaeger & Tily, 2011; Levy, 2008), and has been shown to correlate with cognitive effort during language comprehension (e.g., Hale, 2001; Sikos et al., 2017).

Compared to surprisal, entropy reduction is more directly aligned with the concept of informativity adopted in the current project, as it links each property word to its visually-situated context.

In Information Theory (Shannon, 1948), entropy quantifies the average uncertainty across a set of possible outcomes, based on their probabilities.

$$H = -\sum_{i=1}^{n} P_i log_2 P_i$$

Applying to the referential communication task, entropy represents the uncertainty about which referent is being described in a visual display. Before any

linguistic input, the speaker is assumed to be equally likely to refer to any of the N referents in the context. Thus, the initial probability of referring to any given referent i is:

$$P_i = \frac{1}{N}$$

Applying P_i into Shannon's entropy formula, the initial Entropy is:

$$H_{initial} = -N \cdot (P_i \cdot log_2 P_i) = -N \cdot (\frac{1}{N} \cdot log_2 \frac{1}{N}) = log_2 N$$

For example, the entropy of referring to any of the five objects in Figure 1.1 is $\log_2 5 \approx 2.3$ bits.

Next, when a property word w is used (e.g., "flower"), the set of possible referents matching that property narrows to M items, and entropy is updated accordingly:

$$H_{updated} = log_2 M$$

For "flower", which applies to Object 4 ("the flower-shaped cookie") and Object 5 ("the flower-topped chocolate", as one of possible expressions), $M = 2$, and $H_{\text{updated}} = \log_2 2 = 1$ bit.

After knowing the initial entropy $H_{initial}$, and the updated entropy after mentioning "flower" $H_{updated}$, Entropy Reduction ΔH_w quantifies the decrease of entropy from the initial state to the updated state based on the linguistic input, i.e., the property word w:

$$\Delta H_w = H_{initial} - H_{updated} = log_2 N - log_2 M = log_2 \frac{N}{M}$$

So the entropy reduction of the word "flower" is $\Delta H_{flower} = log_2 5 - log_2 2 \approx 1.3$ bits, which captures the reduction of uncertainty in referring to the possible referents, from five referents to two referents.

By comparison, if the word "cookie" is used first for describing Object 4 (as in "the cookie with the flower shape"), entropy reduced by the word "cookie" ΔH_{cookie} is $log_2 5 - log_2 4 \approx 0.3$ bits, as it applies to four referents ($M = 4$).

Therefore, in the visual context of Figure 1.1, the word "flower" reduces entropy more than "cookie" and is thus more informative.

The current project refers to this measure of informativity as **Referential Entropy Reduction** (RER, Tourtouri et al., 2019), which directly relates each property word to its referential context, i.e., the visual display. Consistent with the qualitative definition of informativity as discriminatory power, RER is related to a property word's **selection scope**, that is, the number of referents that can be modified by the corresponding word. The smaller the selection scope, the higher the RER. For example, in Figure 1.1, "flower" applies to two objects ($RER_{flower} \approx 1.3$ bits), while "cookie" applies to four ($RER_{cookie} \approx 0.3$ bits), making "flower" more informative than "cookie" in this visual context.

RER for each word can be further combined to form a distribution of entropy reduction across multiple property words in a certain linear order, capturing how RER depends not only on the visual context but also on the preceding linguistic context within referential expressions. Building on the idea of selection scope, each successive word in an expression may reduce entropy relative to the set of referents filtered by the preceding words. For instance, describing Object 4 in Figure 1.1 as "the flower-shaped cookie" involves RER of two property words: "flower" narrows the referent set from five to two objects ($RER \approx 1.3$ bits), and "cookie" further reduces it from two to one ($RER = 1$ bit). In contrast, reversing the order, as in "the cookie with a flower shape", yields a different RER distribution: "cookie" first narrows the selection scope from five to four referents ($RER \approx 0.3$ bits), and "flower" subsequently reduces it from four to one ($RER = 2$ bits).

Although both expressions ultimately lead to full certainty in identifying Object 4, the distribution of RER across the expression differs depending on the order in which properties are introduced. This illustrates that the RER of a property word is dependent on both the visual context and the preceding linguistic context. In "the flower-shaped cookie", the more informative word "flower" appears earlier in the utterance, while in the post-nominal alternative expression, "the cookie with a flower shape", it appears later. The current project particularly focused on the choice of the first property word in an expression, as its RER may play a critical role in deciding speakers' choices about property ordering in referential production.

To summarize, both qualitative and quantitative approaches to informativity converge on the idea that a property word is more informative when it distinguishes the target referent more effectively from the rest of the visual context. The present project adopts the quantitative definition of informativity as RER, which captures how much uncertainty is reduced by a given property word in visually-situated contexts and the preceding linguistic contexts.

A final point regarding the notion of "being informative" is that it can also be used to describe the trait of speakers (e.g., Davies & Arnold, 2019; Davies & Katsos, 2010; M. C. Frank & Goodman, 2014). In this sense, speakers are considered informative when they take informativity into account during language production, habitually preferring terms with higher informativity values (e.g., Grodner & Sedivy, 2011; Yildirim et al., 2016). For example, they may strategically select words that refer to smaller selection scopes, thereby facilitating efficient target identification (e.g., Tourtouri et al., 2019). Within computational frameworks such as the Rational Speech Act (RSA) model (M. C. Frank & Goodman, 2012), which will be further introduced in Chapter 2.2.3, informative speakers are "pragmatic speakers" (see also Peloquin et al., 2020), who aim for maximal informativity when only choosing one property word to guide the listener toward the intended referent.

To explore whether and how informativity affects speakers' linearization strategies, it is also important to consider how referential expressions are planned and produced by speakers. The following section provides an overview of linearization during referential production, situating it within broader language production processes shaped by both speaker-internal and speaker-external factors.

1.3 Referential Production and Linearization

In general, referential production is shaped by two broad classes of influences: speaker-internal and speaker-external factors (e.g., Arnold, 2008; Ferreira & Dell, 2000; Gann & Barr, 2012; Wardlow Lane & Ferreira, 2008). Speaker-internal factors involve the cognitive and mechanistic processes underlying language planning, formulation, and articulation (e.g., Bock & Levelt, 1994; Levelt, 1989, outlined in Section 1.3.1 below). In contrast, speaker-external factors focus on aspects of communication contexts, including features of communication environments, interaction dynamics, and assumptions about the identity, perspective, and knowledge state of the addressee (e.g., Fussell & Krauss, 1989; Galati & Brennan, 2010; Heller et al., 2012). Section 1.3.2 introduces these factors briefly, with a more in-depth discussion to follow in Chapter 2.

The current thesis treats informativity and RER as primarily speaker-external factors, as their definitions are grounded in how a linguistic expression helps the listener in identifying the intended referent. While this point will be further elaborated in Section 1.3.2.2 and Chapter 2, it is important to acknowledge that both speaker-internal and speaker-external factors contribute to speakers' choices of property selection and ordering. To situate linearization within the broader process of language production, this section begins with a brief overview of how referential expressions are planned and formulated.

1.3.1 Speaker-Internal Processing of Language Production

According to the language production model proposed by Bock & Levelt (1994) and Levelt (1989), the speaker-internal process of language production consists of three main stages that shape both what speakers say and how they say it: message encoding, grammatical encoding, and phonological encoding (depicted in Figure 1.3, see also Bock & Levelt, 1994).

The **message encoding** stage, also known as conceptualization, prepares the pre-verbal representations for the language formulation process. Examples of pre-verbal representations include, e.g., a conceptual idea, world knowledge, or, in the context of the referential communication task, the visual input about the target referent in a display.

Next, the **grammatical encoding** stage converts these messages into language. It comprises two sub-processes: the functional process and the positional process. During the **functional process**, speakers assign lemmas (e.g., the concept COOKIE) and corresponding lexemes (e.g., specific forms such as "cookie" or "cookies") to the pre-verbal message. In the **positional process**, these lexemes are assembled into a sequence that conforms to grammatical rules, known as constituent assembly in Figure 1.3, determining the linear ordering of words.

Finally, in the **phonological encoding** stage, the structured utterance is mapped onto a phonetic plan for articulation.

Although these stages are often presented as sequential in Figure 1.3, the overall process of production is not strictly linear. It is subject to dynamic adjustments

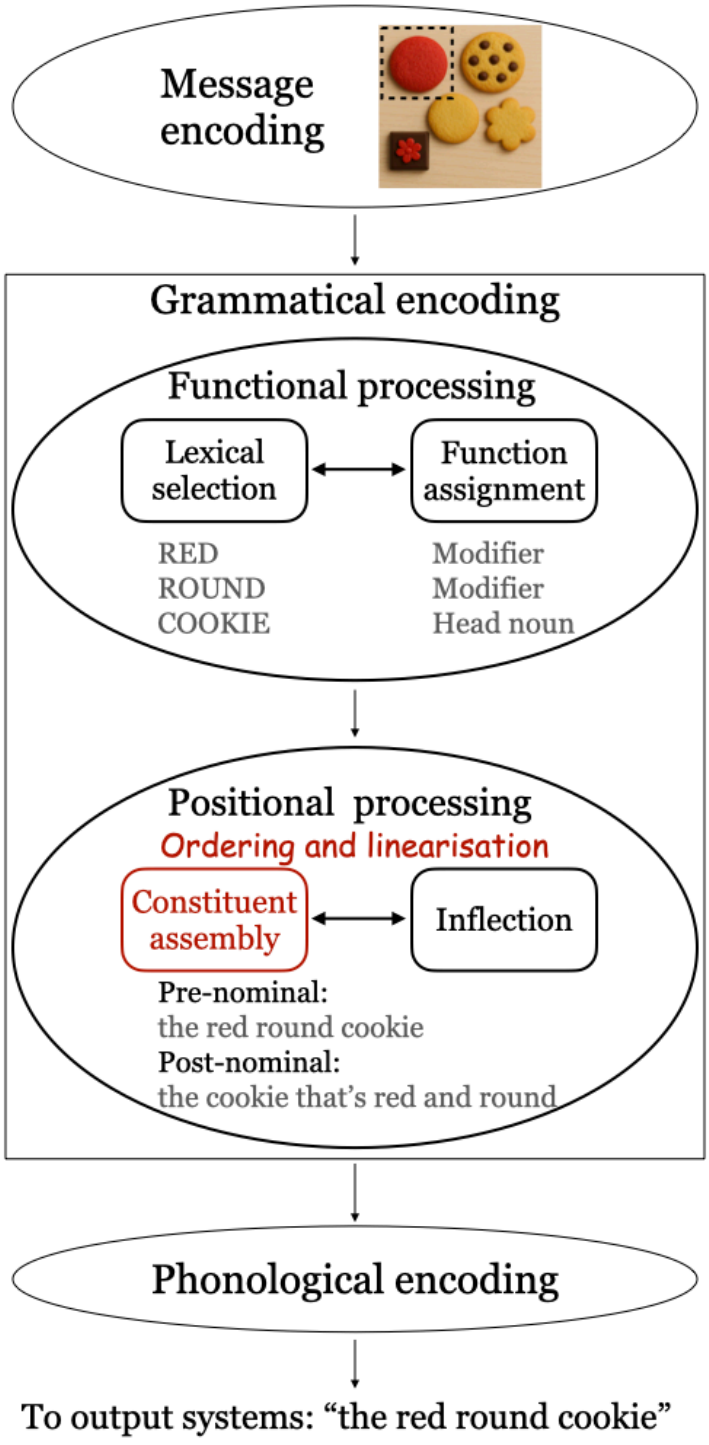


Figure 1.3: An overview of language production processes based on Bock & Levelt (1994) and Levelt (1989), recreated by the author. In this model, linearization is situated at the stage of constituent assembly within positional processing. The dashed square in the visual display marks the target to be described in the message encoding phase. The illustrated linearization contrast focuses on pre-nominal versus post-nominal modifications; other ordering possibilities are discussed in the main text.

through monitoring mechanisms (e.g., Hartsuiker & Kolk, 2001; Postma, 2000) and feedback systems, such as error detection and correction during speech (e.g., Dell, 1986; Levelt et al., 1991).

In the context of referential communication tasks, the production process can be further illustrated as follows. The initial message encoding stage is relatively fixed: the content of the message is constrained by the visual input, namely, the target referent in the display.

During the grammatical encoding stage, the functional process involves selecting the lemmas corresponding to the target's visual properties and assigning their grammatical roles, forming the *modifiers* and the *head noun.* For example, when describing Object 1 in Figure 1.1, possible lemmas include RED, ROUND, and COOKIE. These are realized as the lexemes "red" and "round" (modifiers) and "cookie" (head noun).

The positional process then determines how these selected elements are linearized into a noun phrase (NP). This includes decisions about the syntactic structure, whether modifiers appear in the pre-nominal or post-nominal position (e.g., "the red cookie" vs. "the cookie that's red"), as well as the ordering of multiple modifiers (e.g., "the red round cookie" vs. "the round red cookie"). These two decisions, syntactic linearization and modifier ordering, can also interact, producing expressions such as "the red cookie that's round" or "the round cookie that's red."

Linearization, within the framework of language production depicted in Figure 1.3, takes place specifically during the positional process, where linguistic units are arranged relative to each other into a linear oder based on syntactic, semantic, and potentially pragmatic constraints.

The speaker-internal processing of language production is shaped by a wide range of factors, including cognitive, conceptual, and linguistic influences. Cognitive factors include attention dynamics, for example, speakers are more likely to encode referents that capture their attention as sentence subjects (e.g., Gleitman et al., 2007; Tomlin & Myachykov, 2015), as well as mechanisms such as priming, the tendency to reuse recently activated words or structures (e.g., Cleland & Pickering, 2003; Hartsuiker, 1999; MacDonald, 2013; Pickering & Garrod, 2004). Cognitive effort also plays a role, as speakers tend to prefer utterances that are easy to retrieve or articulate (e.g., Bock, 1986; Kecskes, 2011; Koranda et al., 2022; MacDonald, 2013). Conceptual factors relate to semantic features of the message content, such as a preference for the animate role to appear in subject position (e.g., Tanaka et al., 2011). Linguistic factors include e.g., grammatical constraints and word frequency, both of which can influence production choices (e.g., Griffin & Bock, 1998).

In referential production, relevant speaker-internal influences include visual salience (e.g., color tends to be more salient than size), conventional constraints for modifier ordering (e.g., size typically precedes color in English adjective ordering), and structural priming (e.g., Cleland & Pickering, 2003; Tarenskeen et al., 2015). While the present thesis focuses on the effect of informativity, speaker-internal factors remain important, as they may interact with — or even override — the influence of informativity (e.g., Gatt et al., 2013). Their potential confounding effects were carefully considered and controlled throughout the experimental project of the thesis and will be further discussed in Chapter 2.2.

1.3.2 Speaker-External Influences during Communication

The speaker-external perspective considers language production as situated within a broader communication system that typically involves other interlocutors. Rather than focusing on the internal mechanisms of language production, this perspective examines how speakers' utterances are shaped by their communication functions (e.g., Wardlow Lane & Ferreira, 2008). Speaker-external factors include a range of dimensions regarding communication, such as communication efficiency (e.g., Rubio-Fernández et al., 2021; Rubio-Fernández & Jara-Ettinger, 2020), communication tasks and goals (e.g., Yoon et al., 2012), and sensitivity to the listener's needs or perspective, which is commonly referred to as *audience design* (Clark & Murphy, 1982) or *allocentricity* (e.g., Broisson & Degand, 2022; Peña et al., 2023; Triandis et al., 1985).

These speaker-external, communication-based factors will be further discussed in Chapter 2, especially in Section 2.2.2.4. Two key points motivating the current thesis are introduced in this chapter: First, it remains unclear whether and to what extent speaker-external communication pressures influence linearization, i.e., the grammatical encoding stage of referential production, particularly when they compete with speaker-internal influences in tasks involving communication. Second, the current project conceptualizes informativity as a speaker-external factor in the pursuit of communication efficiency. Informative speakers are considered listener-oriented and engage in audience design when formulating referential expressions.

1.3.2.1 Do Speaker-External Factors Affect Linearization?

To date, it remains debated whether, and to what extent, speaker-external factors influence speakers' initial production decisions. Much previous research has suggested that communication pressures exerted by speaker-external factors are less robust than speaker-internal influences in shaping utterances.

A prominent line of work focuses on *ambiguity avoidance*, where speakers are expected to avoid producing ambiguous expressions that might confuse the listener. However, studies have shown that speakers often fail to avoid lexical or structural ambiguities. For example, speakers tend to refer to a target as "the bat" in a visual context that includes both an animal bat and a baseball bat, thereby failing to uniquely specify the intended referent (Ferreira et al., 2005).

Failures in ambiguity avoidance are especially evident in syntactic choices involving linearization, where speakers do not systematically avoid constructions holding local ambiguity. For instance, Arnold et al. (2004) found that speakers do not avoid ambiguous Prepositional Object constructions (e.g., "Give the letter *to Kim* to me"), although the alternative Double Object constructions ("Give me the letter to Kim") are clearer without local ambiguity. Ferreira (2008) and Ferreira & Dell (2000) showed that the inclusion or omission of the complementizer "that" is primarily driven by availability, a speaker-internal factor, rather than by reducing ambiguity to ease comprehension. Morgan & Ferreira (2022), focusing on the production of resumptive pronouns (further introduced in Chapter 6.3.2.2), concluded that there is little direct evidence so far for audience design effects on syntactic choices.

Moreover, in director-matcher tasks where interlocutors do not view identical visual displays in referential communication, speakers often include information that is only accessible from their own perspectives, even when it is irrelevant or unavailable to the addressee (e.g., Damen et al., 2019; Wardlow Lane et al., 2006; Wardlow Lane & Ferreira, 2008; Wardlow Lane & Liersch, 2012, see Chapter 2.2.2.4). This suggests that utterance planning is often guided more by speaker-internal factors such as availability and production ease (e.g., MacDonald, 2013; Pechmann, 1989) rather than by listener-oriented goals.

However, this does not mean speaker-external factors have no influence on language production. In fact, studies examining spontaneous or naturalistic utterances have revealed robust effects of speaker-external factors on the early message encoding stage of production. For instance, speakers adjust their message content depending on the identity and status of the addressee, for example, when talking to children versus adults (Snow, 1972), or when addressing friends versus strangers (Fussell & Krauss, 1989, see Chapter 2.2.2.4 for more examples). At the phonological level, speakers adjust prosody, such as raising baseline pitch when addressing infants (Kitamura & Burnham, 2003). From information-theoretic perspectives, such as the Uniform Information Density (UID) hypothesis (e.g., A. F. Frank & Jaeger, 2008; Jaeger, 2010; Jaeger & Levy, 2006), corpus studies have shown that speakers may include or omit optional elements (e.g., the complementizer "that") to maintain a uniform distribution of surprisal across the utterance, thereby reducing the listener's processing effort (e.g., Sikos et al., 2017).

In sum, evidence for the effect of speaker-external factors on language production is largely mixed so far. Although speaker-external factors can shape utterance contents at the message encoding level, particularly in spontaneous, unlimited speech, whether the effects can extend to syntactic choices and linearization during the grammatical encoding stage requires further investigation.

The current thesis builds on this debate by investigating whether and how informativity influences the linearization of speakers' referential expressions in support of efficient target identification by the listener — an essential goal of communication efficiency in referential communication tasks (e.g., Fukumura, 2018; Rubio-Fernández, 2016; Rubio-Fernández & Jara-Ettinger, 2020; Tourtouri et al., 2019). Both informativity and the broader goal of communication efficiency are treated as speaker-external factors. The next section explains the reasoning behind this.

1.3.2.2 Informativity as a Speaker-External Factor for Communication Efficiency

Informativity, as defined through RER in the current thesis, refers to the extent to which a property word in a referential expression helps the listener distinguish the intended referent from other objects by progressively narrowing the selection scope in the visual context. This definition suggests that being informative is a listener-oriented goal during referential communication and thus informativity is a speaker-external factor, which reflects how useful a linguistic unit is for the listener's target identification task, rather than how easily it can be produced by the speaker. In this sense, informativity is distinct from speaker-internal factors such as availability (e.g., Ferreira & Dell, 2000) or visual salience (e.g., Gatt et al., 2013; Westerbeek

et al., 2015), which reflect production-based constraints and cognitive preferences (see Chapter 2). Instead, informativity is evaluated in terms of its contribution to communication success and efficiency from the listener's perspective.

Informativity is further closely tied to the goal of communication efficiency. An informative expression facilitates efficient referent identification by the listener (e.g., Fukumura, 2018; Rubio-Fernández, 2016; Rubio-Fernández et al., 2021; Tourtouri et al., 2019). Previous research suggests that speakers are sensitive to such communication pressures: for example, Tourtouri et al. (2019) found that speakers are more likely to overspecify properties that are informative with higher RER in the visual scenes, and Fukumura (2018) showed that speakers tend to prioritize the more informative adjective earlier in referential expressions. These findings indicate that speakers do consider listeners' need for efficient target search during referential production.

The current thesis builds on this perspective by treating informativity as a speaker-external factor that contributes to communication efficiency. Specifically, it examines whether speakers take into account the RER of each property word when determining the order of properties in referential expressions. It focuses particularly on syntactic linearization, that is, how modifiers are sequenced relative to the noun (i.e., in pre-nominal or post-nominal constructions), and asks whether speaker-external factors like informativity influence linearization, achieved during the grammatical encoding phase of language production. This extends previous work on property selection and overspecification (e.g., Tourtouri et al., 2019) and modifier ordering (e.g., Fukumura, 2018) to explore whether and how syntactic variation in property ordering may also reflect listener-oriented strategies of referential production.

Taken together, this section briefly introduced mixed evidence regarding the role of speaker-external factors in language production and motivated informativity as a speaker-external, listener-oriented factor grounded in communication efficiency. By examining how informativity shapes the ordering and structure of referential expressions, the current project investigates whether speakers adjust property ordering according to RER for the communication goal of facilitating listeners' target identification.

1.4 Thesis Overview

The primary objective of this thesis is to investigate whether and how speakers linearize properties in referential expressions to facilitate efficient target identification. Specifically, it examines the role of informativity, quantified as Referential Entropy Reduction (RER), in influencing syntactic ordering decisions — particularly the choice between pre-nominal and post-nominal modification structures.

1.4.1 Research Questions

To address this overarching goal, the following three research questions (RQs) are formulated:

> **RQ1: Does informativity influence property ordering at the syntactic level in referential production?**

The first research question asks whether informativity influences the ordering of properties in referential expressions, particularly when those choices lead to syntactic alternations (e.g., pre-nominal vs. post-nominal structures). Although a few studies have suggested that informativity affects adjective ordering (Fukumura, 2018; Haywood et al., 2003), and its role has been emphasized in computational models grounded in information-theoretic frameworks (e.g., M. C. Frank & Goodman, 2012), no previous research has directly examined whether informativity shapes syntactic ordering choices in referential expressions.

The primary hypothesis is that informativity does play a role in property ordering, particularly in alternating the use of pre- and post-nominal modifications. If speakers take into account the RER of individual properties when choosing between pre- and post-nominal syntactic alternatives, this would indicate that they are not only sensitive to the information-theoretic structure of the visual context but also strategically use it when formulating referential expressions, particularly in property ordering. If any systematic pattern of property ordering modulated by informativity is found, it would reflect a **listener-oriented** approach to referential production, in which speakers aim to support efficient target identification and fulfill broader goals of pursuing communication efficiency.

The alternative hypothesis is that property ordering, particularly at the level of syntactic structure, is not influenced by informativity but is instead governed by speaker-internal factors. These factors, which will be further introduced in Chapter 2.2, include e.g., grammatical and semantic constraints (e.g., "the cookie that is red" for Object 1 in Figure 1.1 would be rare for English), the preference to mention visually salient properties first (e.g., color, Fukumura, 2018; Gatt et al., 2013), as well as effects of structural priming and alignment (e.g., Pickering & Garrod, 2004; Tarenskeen et al., 2015).

These speaker-internal factors may lead speakers to rely on the preferred syntactic structure[4] for referential encodings regardless of informativity. If such consistent patterns are observed repeatedly across different experimental tasks and communication settings, it would suggest that informativity does not influence linearization, supporting a **speaker-oriented**, heuristic production strategy, where

[4]In this thesis, I use the term *preferred* syntactic structure or *preferred* word order to refer to forms that speakers tend to use most frequently in neutral or unmarked contexts, when there is a clear preference among alternative expressions (e.g., preferring the pre-nominal structure such as "the blue dress" rather than the post-nominal structure as in "the dress that is blue" in English). The term can also refer to the habitual choice of an individual speaker: for example, a speaker who consistently favors post-nominal structures, such as "the dress that is blue", regardless of the underlying reason, may be said to have an individual preference for that form.

communication efficiency and informativity, as speaker-external factors, do not affect the grammatical encoding stage of reference production.

The comparison between the speaker-oriented and listener-oriented views, which can be clearly teased apart by the presence or absence of the effect of informativity, will be further introduced in Chapter 2.

RQ2. If informativity does influence property ordering, how does it linearize property orders specifically?

If informativity affects the choice between pre- and post-nominal modifications, the second research question concerns the specific pattern of this influence. The main hypothesis is that speakers adopt an **informative-first linearization preference** (in line with e.g., Cohn-Gordon et al., 2019; Fukumura, 2018; Haywood et al., 2003, elaborated in Chapter 2.2), placing the more informative property word (i.e., the one with higher RER) earlier in the referential expression. This prediction is supported by previous findings showing that more informative properties can significantly narrow the selection scope of referential candidates, reducing uncertainty about referring to the target referent (Fukumura, 2018; Tourtouri et al., 2019), and thus facilitate the listener's visual search incrementally during processing (e.g., Rubio-Fernández et al., 2021).

A possible alternative pattern is motivated by the Uniform Information Density (UID) hypothesis (e.g., A. F. Frank & Jaeger, 2008; Jaeger, 2010), which would predict that speakers prefer utterances that distribute informativity more uniformly across an expression. This view predicts an ordering pattern opposite to the informative-first linearization hypothesis: speakers might place the more informative word later in the expression to avoid an early peak of informativity, thereby reducing the listener's processing cost.

However, this UID-based prediction remains only exploratory in the current thesis, as several theoretical assumptions are required to extend UID to the present context. First, it requires the assumption that RER functions analogously to surprisal as a measure of informativity (see Section 1.2.2), to satisfy the UID hypothesis that among alternative encodings, speakers prefer to choose the one that can distribute information density more uniformly across linguistic signals to avoid peaks of surprisal, if extending to the current case, peaks of RER. Second, it presumes that RER correlates with cognitive effort in visual search, in the same way as how surprisal is correlated with processing effort in language comprehension, where words with high surprisal peak requires excessive comprehension effort for readers (Hale, 2001; Sikos et al., 2017; but see also Tourtouri et al., 2019 for an initial attempt to link RER with processing cost). Third, applying UID to referential expressions requires extending the notion of uniform information distribution beyond syntactic reductions and omissions of function words (e.g., Jaeger, 2010; Jaeger & Levy, 2006) to ordering choices among content words. Finally, one must also assume that UID applies to *restrictive* linguistic uses (Hahn et al., 2018), such as referential expressions constraint by visual contexts, although most UID evidence stemming from *non-restrictive* utterances in open-ended discourse, such as telephone dialogues (Jaeger, 2010), where the message encoding phase of language production is not limited in the first place.

Because these assumptions remain largely theoretical, the UID-based prediction of *informative-later* ordering is treated as an exploratory contrast. The informative-first linearization hypothesis remains the central prediction of this thesis. RQ1 and RQ2 are mainly explored in Experiments 1-3 reported in Chapter 4.

RQ3. Does communication engagement further modulate the effect of informativity on property ordering?

The final question investigates whether **communication engagement** influences the extent to which informativity affects syntactic linearization. Previous research suggests that communication environments and task experiences can influence speakers' level of engagement and the chance to adopt the listener's perspective, such as when comparing human-to-computer versus human-to-human interaction (Peña et al., 2023), or tasks with varying degrees of interaction and role switching (e.g., Vogels et al., 2020). These differences may, in turn, influence how speakers consider the listener's needs and how they integrate informativity into the ordering of properties.

To test this, the current project systematically varies communication engagement across experiments: from less interactive, online communication tasks to highly engaging face-to-face dialogues in which participants alternate between speaker and listener roles. If more engaging environments indeed amplify the effect of informativity on linearization, this would provide further evidence that informativity functions as a speaker-external, communication-based factor on property ordering. It would also suggest that informative speakers are listener-oriented, actively pursuing cooperative and efficient communication by incorporating informativity into the planning and formulation of referential expressions. RQ3 is primarily addressed in Chapter 5.

1.4.2 Thesis Structure

The thesis is structured as follows:

Chapter 2 provides a literature review of previous research on property selection (Section 2.1) and property ordering (Section 2.2). This chapter first argues that informativity plays a central role in distinguishing between the speaker-oriented view (Section 2.1.1) and the listener-oriented view (Section 2.1.2) of overspecification. It then proposes shifting the focus from studying property selection to property ordering, as property ordering may offer clearer insights for differentiating these views. Finally, it reviews factors influencing property ordering, especially psycholinguistic factors which are crucial for motivating the current project (Section 2.2.2). An overview of the experimental project will be provided at the end of this chapter (Section 2.3).

Chapter 3 presents two pre-tests conducted to validate the experimental materials and linguistic structures used in the project. These pre-tests ensured that the visual stimuli (animals performing actions, e.g., "a crying rabbit") allowed flexible use of pre- and post-nominal modifications in German.

Chapter 4 addresses RQ1 and RQ2. Through three experiments employing a maze-based sentence completion task in an online communication environment, this chapter investigates whether informativity influences property ordering at the syntactic level and characterizes the specific linearization patterns preferred by speakers.

Chapter 5 focuses on RQ3, presenting two experiments that enhanced communication engagement by alternating participants' speaker and listener roles in both online and face-to-face communication settings. These experiments explore how communication dynamics modulate the influence of informativity on linearization.

Chapter 6 discusses the key experimental findings and contributions of the project in the broader context of communication efficiency and language production. It also addresses limitations of the referential communication task paradigm and outlines directions for future research.

With this introduction established, the next chapter reviews previous research on property selection and property ordering in referential production, providing the background for the experimental investigations of the current project.

1.4.3 Related Publications and Presentations

Several experiments from this dissertation have been presented at conferences in various formats:

Conference paper:

- Li, M., Venhuizen, N. J., Jachmann, T. K., Drenhaus, H., & Crocker, M. W. (2023). Does informativity modulate linearization preferences inreference production? *Proceedings of the 45th Annual Conference of the Cognitive Science Society*, *45*, 3048–2054. https://escholarship.org/uc/item/95v6j0sx

Conference Talks/Posters:

- Li, M., Venhuizen, N. J., Jachmann, T. K., Drenhaus, H., & Crocker, M. W. (2023, August 31 - September 2). *Does informativity modulate linearization preferences inreference production?* [Oral presentation]. Architectures and Mechanisms for Language Processing (AMLaP), Donostia–San Sebastián, Spain. https://www.bcbl.eu/events/amlap/en/conference-program/su-107/

- Li, M., Venhuizen, N. J., Jachmann, T. K., Drenhaus, H., & Crocker, M. W. (2023, September 20-22). *Does informativity modulate linearization preferences inreference production?* [Poster presentation]. The Tenth Biennial Meeting of Experimental Pragmatics (XPrag), Paris, France. https://xpragx.sciencesconf.org/

- Li, M., Venhuizen, N. J., Jachmann, T. K., Drenhaus, H., & Crocker, M. W. (2025, February 13-15). *Informativity modulates linearization preferences in referential interaction* [Poster presentation]. Rational Approaches in Language Science (RAILS), Saarbrücken, Germany. https://sfb1102.uni-saarland.de/sfb-conference-2025/rails-2025-program/

2

Informative Property Selection and Property Ordering in Referential Communication

Contents

Building on the central research question of this thesis — how informativity affects the linearization of referential expressions — this chapter reviews previous research on the role of informativity in two key processes involved in referential production: 1) *property selection*, the choice of which properties speakers include to describe a target referent, and 2) *property ordering*, the sequencing of those properties into linear syntactic form. While informativity has been shown to guide property selection, particularly in relation to overspecification, its role in property ordering remains less well understood.

The chapter begins by examining studies on property selection in referential communication tasks, with a particular focus on overspecification (OS) and the conditions under which it commonly occurs. Two contrasting theoretical accounts, the speaker-oriented and the listener-oriented view, are discussed to explain why

OS arises, with informativity serving as a key factor in distinguishing between them. OS is introduced first in this chapter because insights from studies regarding OS are also relevant to research on property ordering, and thus to the current project. The first half of the chapter concludes with reflections on the limitations of using OS as a primary measure of referential production, motivating a shift in focus toward linearization as the central object of the current study.

The second half of the chapter turns to property ordering in referential production. This section examines property ordering from three perspectives: conventional constraints, psycholinguistic factors, and information-theoretic accounts. Although most existing studies have focused on the ordering of pre-nominal adjective modifiers, relatively little work has investigated post-nominal structures or property ordering that results in syntactic variation.

The chapter concludes with an overview of the experimental project presented in the following chapters, based on theoretical insights developed throughout this chapter regarding both property selection and property ordering.

2.1 Informative Property Selection in Reference Production

Property selection — also referred to as *content selection* (e.g., Gatt et al., 2013; Koolen, 2019) — has been extensively studied by examining how properties encoded in referential expressions correspond to those necessary for uniquely identifying a target in visually-situated contexts. As introduced in Chapter 1, this has led to categorizing referential expressions into three types: *underspecification*, *minimal specification* (MS), and *overspecification* (OS). While MS ideally aligns with the Gricean Maxim of Quantity (Grice, 1975), previous research has shown that OS is strikingly common: speakers frequently include more information than necessary when describing a target, particularly under the following scenarios:

1. OS is frequently adopted when the corresponding visual properties are salient and absolute, such as color or pattern (e.g., "the *blue* ball" in Tourtouri et al., 2019; "the *striped* socks" in Tarenskeen et al., 2015), or when referring to small cardinalities (e.g., "*two* stars" in S. A. Wu & Gibson, 2021; Zevakhina et al., 2021). OS is less frequently adopted for properties that are relative and context-dependent, requiring mandatory comparison with other objects, such as for size (e.g., "the *small* butterfly" in Brown-Schmidt & Konopka, 2011), or when the property is difficult to perceive, such as for material (e.g., "the *metal* chair" in Kursat & Degen, 2021).

2. OS is frequently adopted when the visual scene is complicated, in the sense that it involves multiple dimensions of object properties (e.g., as in Koolen et al., 2013, which included four property dimensions in the stimuli: color, size, orientation, and type). Higher visual complexity can also result from an increased number of objects in the scene (Elsner et al., 2018). It may also more likely to arise when the scene is polychrome (featuring multiple

colors) rather than monochrome (with only one color for all objects, Rubio-Fernández, 2019), or when the property of the target itself is conceptually complex, for example, describing a real-life portrait such as "an old man with a beard and glasses" (Koolen et al., 2011).

3. OS is frequent when one of the properties of the target is unique and visually pop-out (e.g., the "pop-out" condition in Rubio-Fernández, 2019, where only the target is blue while the remaining objects are yellow; see also Gatt et al., 2017).

4. OS is more likely to occur when the color of the target is atypical (e.g., a *blue* tomato, Sedivy, 2003; Westerbeek et al., 2015), or when color is relevant and important to mention for certain types of target referents (e.g., "a *yellow* banana" vs. "a *yellow* shirt," where the color yellow is more important to mention for a shirt than a banana that is typically yellow, Rohde & Rubio-Fernández, 2022; Rubio-Fernández, 2016).

5. OS is also more likely to occur when the speaker has already produced an overspecified expression in a previous trial, even with a different property. For example, the frequency of size OS increases when color OS has previously been produced by the same speaker (referred to as *the consistency effect* in Tarenskeen et al., 2015).

6. OS is also frequent when the property is obligatorily encoded in a pre-nominal rather than a post-nominal position. For example, color is more frequently overspecified in English, where it appears as a pre-nominal adjective (e.g., "the *blue* dress"), compared to Spanish, where it appears post-nominally (e.g., "el vestido *azul*" [the dress blue], Rubio-Fernández, 2016; S. A. Wu & Gibson, 2021).

Although the conditions under which OS arises, as listed above, have been extensively studied, there is less consensus on the underlying reasons for its occurrence. A common starting point for explaining the frequent use of OS is that it emerges within the referential communication task (see Chapter 1.1), where a speaker conveys a message to a listener within a shared visual environment (Clark & Wilkes-Gibbs, 1986). Understanding OS, then, must be tied back to this communication process. Accordingly, researchers have asked whether OS provides any benefit, and if so, for whom in communication: the speaker or the listener? This question has given rise to two theoretically competing perspectives: *the speaker-oriented view*, which argues that OS primarily benefits the speaker, and *the listener-oriented view*, which posits that the speaker produces OS to benefit the listener.

Before examining how these two views relate specifically to OS, it is important to recognize that the two views are not only limited to accounting for OS, but have also been applied across a wide range of communication phenomena, such as the use of discourse marker (e.g., Broisson & Degand, 2022), disfluency (e.g., Engelhardt et al., 2017; Lake et al., 2011) and gesture (e.g., Ianì & Bucciarelli, 2017; Özyürek, 2002). To provide a foundation for their application to OS, the two views are briefly outlined here at a conceptual level.

The **speaker-oriented view**, which is an egocentric, heuristic-based perspective (Horton & Keysar, 1996), posits that speakers' utterances are primarily guided by their own cognitive constraints and processing ease. The speaker-oriented view emphasizes the speaker-internal processes, such as memory retrieval or production ease, suggesting that communication and production are driven by what is easy or natural for the speaker (e.g., Ferreira et al., 2005; Ferreira & Dell, 2000; MacDonald, 2013). Routinely considering the listener's perspective, by contrast, would require additional effort and cognitive resources that may exceed the speaker's cognitive capacity (e.g., Keysar et al., 1998).

In contrast, the **listener-oriented view** suggests that speakers actively consider the perspective of the addressee when constructing utterances (known as *audience design* in e.g., Clark, 1992; Clark & Carlson, 1982; Clark & Murphy, 1982; also relevant to the term *allocentricity* in e.g., Broisson & Degand, 2022; Peña et al., 2023; Triandis et al., 1985). The listener-oriented view emphasizes the role of common ground, i.e., the shared knowledge, assumptions, and beliefs between interlocutors (Clark & Marshall, 1981) in shaping speakers' utterances and behavior, potentially even at early stages of language processing (e.g., Brown-Schmidt & Hanna, 2011).

These two views can refer not only to behavioral strategies during language production (e.g., tailoring an utterance to meet the listener's needs of target identification as a listener-oriented strategy) but also to characteristics of individual speakers. For example, some speakers may be described as speaker-oriented or egocentric in their communication style (e.g., Horton & Keysar, 1996; Wardlow Lane & Liersch, 2012), while others may be characterized as listener-oriented or allocentric (e.g., Peña et al., 2023).

This section reviews empirical studies and debates surrounding OS, focusing on the comparison between the speaker-oriented and listener-oriented views. It highlights the role of informativity, suggesting that the divergence between these views may stem from how speakers handle informativity. The section will then discuss whether focusing solely on property selection, particularly OS, is the best case study for this debate. It will argue for a shift to focus on property ordering, i.e., linearization, for more insights. The section will conclude by exploring how certain hypotheses stemming from the study of OS can inform future research on linearization.

2.1.1 Speaker-Oriented View

The speaker-oriented view, as primarily supported by Ruud Koolen and his colleagues (e.g., Gatt et al., 2013; Koolen et al., 2013; Koolen, Krahmer, et al., 2016; Koolen, 2019), proposes that OS is a heuristic strategy of referential production. Heuristics in psychology refers to "mental shortcut" or simple, efficient rules used in problem-solving and decision-making, allowing people to arrive at solutions quickly without the need for complex analysis (Simon & Newell, 1958; see a review Hjeij & Vilks, 2023). In the speaker-oriented view, OS may function as a heuristic strategy for three main reasons:

Figure 2.1: Visual stimuli example from Koolen et al. (2013), Figure 2. The figure illustrates differences in visual complexity based on scene variation: the display on the right contains more property variation and is therefore visually more complex. Reproduced with permission from *Cognitive Science*, © 2013 John Wiley & Sons. License number: 6018771220307 (issued April 30, 2025).

First of all, by using OS, speakers do not have to compare the target with every other referent in a visual scene before speaking, saving the cognitive effort of visual comparison. Eye tracking evidence in Pechmann (1989) showed that speakers already started speaking and frequently used OS before viewing the full visual scene, making reference production an incremental process. Belke & Meyer (2002) proposed that formulating referential expressions involves a series of decisions based on "same/different" verification for each property between the target and the non-target objects, a process they termed the "relevance rechecking model". Producing MS requires speakers to evaluate each property dimension, which demands considerable cognitive effort. In contrast, mentioning an absolute and salient property (e.g., color) without any comparison offers a more straightforward and cognitively efficient strategy for speakers.

The heuristic strategy of producing OS becomes even more beneficial for speakers when the scene involves high visual complexity. A scene is more visually complex not only when it contains a large number of referents in total (e.g., 16 objects in Rubio-Fernández, 2019; 121 objects in Elsner et al., 2018; adding more objects as "visual clutters" in Koolen, Krahmer, et al., 2016), but also when more dimensions of object properties exhibit variations. For example, Koolen et al. (2013) designed a series of visual stimuli similar to Figure 2.1 where the display on the right exhibits greater scene variation than the display on the left, as it varies in three dimensions of object properties: color, object orientation (e.g., facing left or right) and object type (e.g., sofa or TV), whereas the left display varies only in one dimension, i.e., object type. Another case of high visual complexity is when there are more variation within a property dimension, e.g., in Figure 2.1, visual complexity of the right display is higher also because it includes multiple colors (green, grey, brown, blue, red), while the left display contains only one color (green).

Koolen et al. (2013) detected a higher frequency of OS adopted in the high-complexity condition, i.e., speakers would produce more OS when describing the target in the right display in Figure 2.1. They suggested that the increased use of OS serves as a heuristic strategy: navigating complex scenes demands higher

cognitive effort, leading to longer viewing times and extended utterance planning, as evidenced by longer speech onsets (Gatt et al., 2017; Koolen & Krahmer, 2020). When viewing time was limited by the experimental design and speakers lacked the possibility for thorough inspection of the visual scene, the frequency of OS would also rise (Koolen, Gatt, et al., 2016), probably because OS serves as a quick problem-solving strategy under time pressure.

The second reason to interpret OS as a heuristic strategy is that speakers can repeatedly employ this strategy, without the need to assess the potential change of the visual scenes and the state of the listener, while still ensuring communication success for the listener to accurately identify the target. Tarenskeen et al. (2015) observed a consistent pattern in speakers' use of OS: while it has known that OS rates are higher for absolute and salient properties like *color* and *pattern* (with color OS often at ceiling level) compared to relative properties like *size*, the exact OS rates for these properties can vary based on the experimental design. In a within-subject design (Experiment 1), participants encountered trials allowing OS for color, pattern, or size, while in a between-subjects design (Experiment 2), participants only experienced one property type possible for OS across all trials. The results showed that OS for *pattern* was almost as high as *color* in the within-subject design but dropped significantly in the between-subject design. These findings suggest that once speakers adopt OS for one property, they are likely to apply it to other properties repeatedly.

The consistency trend of OS has not only been observed across different visual properties, but also across changes of the audience paired with speakers: Gann & Barr (2012) implemented the referential communication paradigm consisting of a training phase and a testing phase. During the training phase, the target was paired with a competitor, requiring the expression to include a modifier (e.g., "an *unmelted* candle", in contrast to a melted candle shown in the same display). In the testing phase, the competitor was replaced with an object that did not share properties with the target. As a result, using a bare noun without the modifier (e.g., "the candle") was sufficient to specify the target presented in the testing phase, and continuing to use the modifier constituted OS (e.g., while the target remained to be "the unmelted candle," but there was only one candle in the display, making the use of "unmelted" an instance of OS).

The key manipulation of this study was that listeners paired with half of the participants were replaced with a new group of listeners who encountered the display and the target for the first time. The question was whether speakers would maintain or reduce the use of the overspecified modifiers for these new listeners. The results showed a consistent pattern of OS across both phases: speakers continued to prefer OS even though the overspecified modifier was not necessary for the new listeners to identify the target. This finding suggests that OS is a speaker-oriented decision, as a listener-oriented speaker would have used a bare noun that was sufficient for the new listener group.

Gann & Barr (2012) argued that, at least for conventional objects and familiar targets, the high rate of OS should be driven by speakers' prior experience and memory from the training phase, rather than the updated status of the audience or the visual contexts. Notably, speakers did not rely solely on heuristic strategies

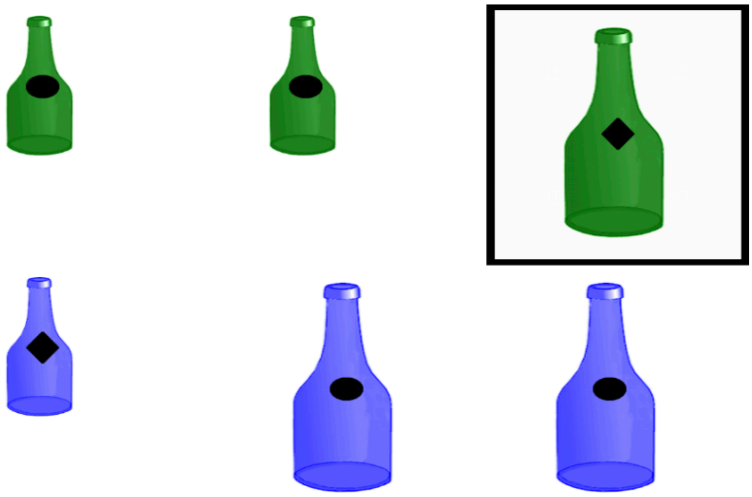


Figure 2.2: Example stimulus in Gatt et al. (2013), Figure 1. Including any two of the properties minimally specifies the target, while including all three constitutes OS. Reproduced from the original under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

in this study: when the targets were unconventional objects that could not be easily named with a noun (e.g., "stacked snail shapes"), speakers used more words to describe these targets, particularly when listeners had not encountered them before, which, on the other hand, supported the listener-oriented view (discussed in the following Section 2.1.2).

The consistency manner of OS in Tarenskeen et al. (2015) and Gann & Barr (2012) both indicated that OS can be a heuristic strategy for speakers because they can repeatedly adopt the same OS pattern in front of changes of visual contexts (e.g., across different visual properties, or when the competitor of the target disappeared) as well as changes of the addressees. More importantly, although OS may risk impeding comprehension, as evidenced by slower reaction times in target searches by the new listener group after hearing OS in Gann & Barr (2012), the overall communication goal of accurately identifying the target can still be successfully achieved through the use of OS as a heuristic.

The final reason to treat OS as a heuristic strategy is that informativity may not always be a primary consideration for speakers. Considering informativity in referential production requires linking property words to the visual context. However, this link may not yet be well-established before speaking if speakers rely on heuristics and do not engage in thorough visual comparison between the target and other referents. Other speaker-internal factors, such as the preference for color due to its high visual salience (e.g., Clarke et al., 2013; Wardlow Lane & Ferreira, 2008; Westerbeek et al., 2015), may play a bigger role.

Evidence was observed in Gatt et al. (2013), which suggested that discriminability (i.e., informativity) might not influence property selection. In the study, speakers were asked to describe a target in a visual context such as Figure 2.2, where

any two properties out of the three, *color*, *size*, and *shape inside the bottle*, could be used to minimally specify the target (e.g., "the green bottle with a diamond," "the large green bottle," or "the large bottle with a diamond"). Mentioning all three properties would otherwise form OS. Across three conditions, each property (*color*, *size*, and *shape*) was the most discriminating in turn, meaning it narrowed the selection scope to two objects, while the other two properties narrowed the scope to three (e.g., in Figure 2.2, *shape* diamond is the most discriminating).

The key hypothesis was that if discriminability influenced property selection, speakers would be more likely to use minimally specified expressions that included the most discriminating property. However, if discriminability had no effect, speakers would choose properties based on heuristic preferences, such as visual salience (favoring *color* over *size*) or syntactic structure (favoring pre-nominal over post-nominal modifications, i.e., color and size over shape).

The results showed that discriminability did not influence property selection: the rates of MS and OS were consistent across all three conditions (approximately 70% MS vs. 30% OS). Within MS, property selection was unaffected by discriminability but was instead driven purely by the preference ranking of the three properties (color > size > shape). Gatt et al. (2013) concluded that discriminability cannot override the effect of property preference and does not play a significant role in property selection.

However, it is important to note that Gatt et al. (2013) cannot conclusively determine the absence of an effect of informativity on OS: the contrast in discriminability in this study was relatively small. A difference of merely one object may not have been sufficient for speakers to plan their expressions based on discriminability. In addition, the mechanism behind property selection, particularly the hierarchy of how each property was chosen, remains unclear and is not fully captured by the reported results. Participants may have initially selected the most informative property and then decided to add others, or they could have chosen properties based purely on preference or the canonical order of the modifiers. Therefore, the findings cannot fully rule out the possibility that informativity did play a role in property selection, but was not captured by the experimental stimuli and design of this study.

To sum up, under the speaker-oriented view, OS is considered as a heuristic strategy in reference production. OS allows speakers to not only reduce the cognitive effort of visual search before speaking, but also repeatedly adopt the same easy approach to guarantee basic communication success regardless of changes in visual contexts and the status of addressees. Informativity, under this view, was not an influential factor in deciding property selection.

2.1.2 Listener-Oriented View

In contrast to the speaker-oriented view, the listener-oriented view argues that OS should not simply serve as a heuristic to facilitate speaking but should also provide benefits for the listener. A key hypothesis within this listener-oriented view, particularly in research on OS, is that OS can facilitate the listener's visual search for the target. A comprehensive understanding of this hypothesis requires

exploring both: 1) whether speakers genuinely tailor their expressions to meet the listener's needs, i.e., whether OS is produced with the listener-oriented intentions, and 2) whether listeners truly benefit from this strategy when visually searching the target referent.

This section will introduce research that sheds light on these two questions. It will begin by addressing the second question — whether listeners truly benefit from OS — by reviewing studies on the comprehension side that measure the listener's reaction times, eye movements, and event-related brain potentials (ERPs). While not conclusive, most previous studies suggest that OS facilitates visual search, particularly when it helps narrow the selection scope for identifying the target, that is, when the overspecified property word is informative (see Chapter 1.2.1). This section will then discuss the first question concerning whether speakers use OS truly to tailor to the needs of listeners, that is, to help listeners identify the target efficiently. The general conclusion is that, although not always, OS are produced with listener-oriented intentions, where informativity plays a key role in determining why certain properties are overspecified.

Another constraint for the studies discussed in this section is that the referential communication should arise from tasks where both interlocutors share identical visual information, as this assumption was also implemented in the experiments of the current thesis project. However, it is important to note that the listener-oriented view in referential communication has also been examined in other contexts: e.g., when interlocutors do not share the same information, with the speaker having more *privileged knowledge* than the listener. This will be further discussed in Section 2.2.2.4.

2.1.2.1 Can Overspecification Facilitate Target Identification?

Can listeners benefit from OS during visual search of the target? Empirical findings are mixed regarding whether OS facilitates or impedes comprehension. Some researchers argue that overspecified information may hinder comprehension, as listeners could assume that the additional detail carries implicit communicative relevance (e.g., Engelhardt et al., 2011) or reflects additional speaker intentions (Grice, 1975). In referential communication tasks, however, such pragmatic implications may be lacking, and the overspecified detail may instead introduce interference that disrupts comprehension. Depending on the specific display, listeners may also draw inferences toward unintended referents by attending to competitors that could plausibly be described by the overspecified information, thereby increasing processing difficulty (Engelhardt et al., 2006).

Support for the hindering effect of OS comes from studies showing slower reaction times when listeners search for targets guided by overspecified expressions. For instance, in the study by Gann & Barr (2012), listeners were approximately 200 ms slower to click on the target when hearing OS (e.g., "the *unmelted* candle") compared to a minimally specified expression (e.g., "the candle"). Similarly, Engelhardt et al. (2011) found longer reaction times when the target was overspecified (e.g., "the *red* star" in a display with only one star) compared to the same phrase but functioning as MS in a different display (e.g., "the red star" in a display with both red and blue stars). Additional support comes from ERP studies: the N400 component,

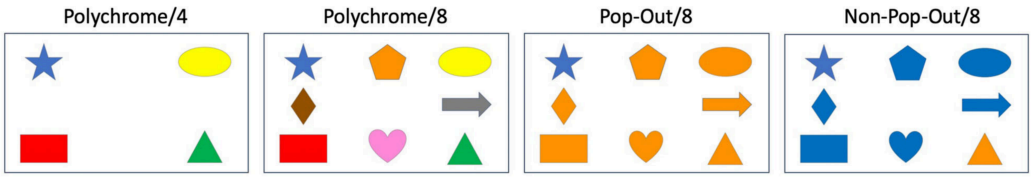


Figure 2.3: Example stimulus from Rubio-Fernández (2021), Figure 3. "The blue star" is the OS of the target. The informativity of the color word "blue" varies across conditions: it is less informative in the Non-Pop-Out/8 condition than in the other three. Reproduced from the original under the terms of CC BY 4.0.

indexing semantic processing difficulty (Kutas & Hillyard, 1980), has been shown to be more negative in response to OS, suggesting increased difficulty in integrating the redundant information into target identification (Engelhardt et al., 2011).

However, other research has found that OS facilitates the listener's visual search for the target. This facilitation effect has been observed across various types of visual scenes. For example, Arts et al. (2011) found shorter target identification times after participants first read OS descriptions of the target, particularly when the overspecified properties contributed to forming a complete mental image of the target, helping to reduce the visual search load. Rehrig et al. (2021), using the eye-tracking technique to examine visual search in real-world scenes (e.g., finding a black lamp in a photo of a laboratory), reported that overspecified color and location information reduced visual scanning time before participants fixated on the target. OS has also been shown to facilitate spatial search: participants' searching times and distances were shorter when guided by OS instructions in a 3D virtual space, where the targets were hidden behind landmarks in separated rooms that were either uniquely identifiable or shared common properties with other referents (Paraboni & van Deemter, 2014).

Unlike the speaker-oriented view, the listener-oriented view places informativity at the core of motivating OS. The reason OS can facilitate the listener's visual search is largely due to the high informativity value of the overspecified message. Research suggests that visual search for the target occurs incrementally: listeners do not wait for an utterance to be complete to initiate visual search, but begin narrowing down potential referents with each incoming word (e.g., Eberhard et al., 1995). When an overspecified property word is more informative, especially if it appears early in an utterance, it significantly reduces the selection scope, thus facilitating the target search process (e.g., Rubio-Fernández et al., 2021; Rubio-Fernández, 2021).

Evidence of the facilitation effect based on informativity has been observed in the listener's reaction times and eye movements. In Experiment 1 from Rubio-Fernández (2021), listeners were presented with visual contexts such as Figure 2.3, and were guided by either MS ("Click on the star") or OS ("Click on the *blue* star"). The four conditions can be interpreted in terms of the informativity of the overspecified word "blue". In the Non-Pop-Out/8 condition, where most referents are blue, "blue" was less informative compared to the other three conditions, where it could directly identify the target. The reaction times and eye-tracking results

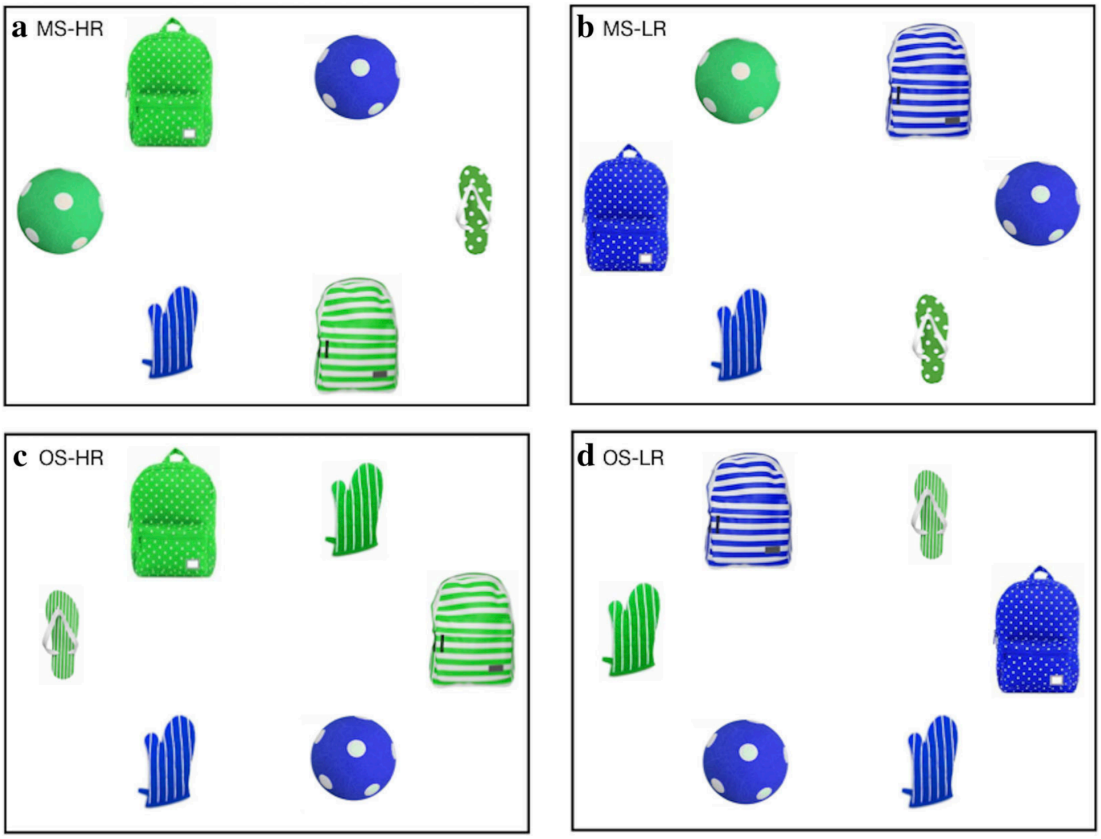


Figure 2.4: Example stimulus in Tourtouri et al. (2019), Figure 1. "The blue dotted ball" is the target. The informativity of the color word "blue" is different across conditions, quantified by High- or Low-Entropy reduction (HR or LR). Reproduced from the original under the terms of CC BY 4.0.

supported the facilitation effect of OS: listeners clicked on the target faster when guided by OS compared to MS in the first three conditions, where "blue" was informative, but slower in the Non-Pop-Out/8 condition. Eye-tracking data further showed that listeners fixated more on the target in an earlier phase when hearing the overspecified adjective ("blue") than when directly hearing the bare noun ("star").

Tourtouri et al. (2019) tested the facilitation effect of OS by quantifying the informativity of the overspecified property word by RER. For instance, in Figure 2.4, while "the blue ball" is MS in Conditions a) and b), it is OS in Conditions c) and d) where there is only one ball in the visual scene. The word "blue" is more informative in Conditions a) and c), as it narrows the reference to two objects (High Entropy Reduction, HR), compared to four objects in Conditions b) and d) (Low Entropy Reduction, LR). Although the target is ultimately identified at the noun ('ball') in all conditions, their results showed that cognitive effort at the noun — indicated by Index of Cognitive Activity (ICA), a pupillometric measure of cognitive load — was reduced when the overspecified adjective is informative (Condition c). They also detected increased fixations on the target in the time window of hearing the informative adjective (e.g., "blue") and shorter reaction times in target identification in this condition. These evidence suggested that informative

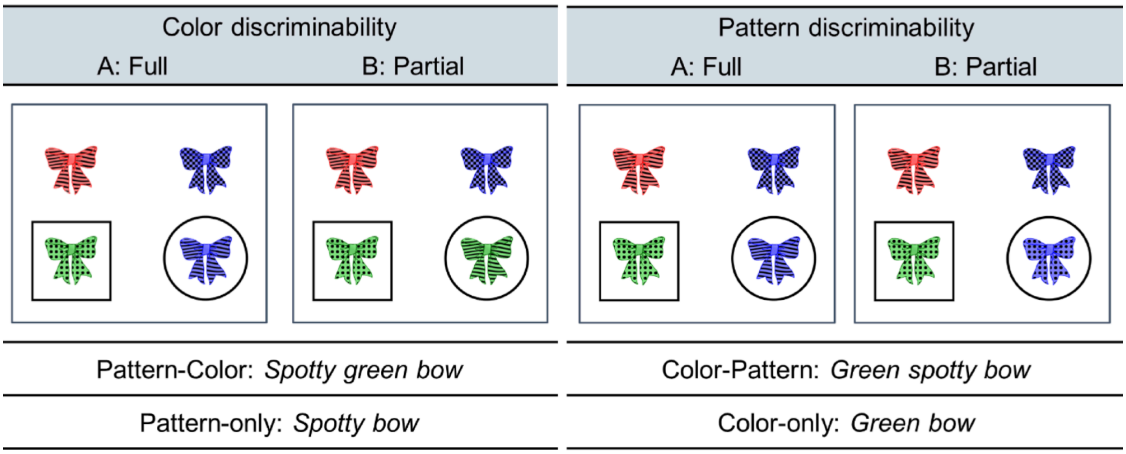


Figure 2.5: Example stimulus in Fukumura & Carminati (2021), Figure 6 (left) and Figure 13 (right). The target referent is shown in a box, and the competitor in a circle. In both experiments, although a "Pattern-only" or "Color-only" expression is sufficient to minimally specify the target, the second adjective in the full expression (Pattern–*Color* or Color–*Pattern*) differs in its discriminability between Condition A (full discriminability) and Condition B (partial discriminability). Reproduced with permission from *Journal of Experimental Psychology: Learning, Memory, and Cognition*, © 2022 American Psychological Association. License number: 6018800001307 (issued April 30, 2025).

OS can facilitate comprehension, especially when the overspecified property word is informative with high RER.

Across a series of experiments varying the discriminability of color and pattern properties, Fukumura & Carminati (2021) suggested that although informativity plays an important role, the facilitation effect of OS is also dependent on which property is overspecified and when it occurs before listeners start the visual search process (see also Mangold & Pobel, 1988). In Experiments 2 and 3 (Color- or Pattern-discriminability in Figure 2.5), they investigated whether the second adjective, which was always overspecified but differing in discriminability, would hinder or facilitate comprehension. They also varied the timing of the presence of a visual display, presenting it either simultaneously or 1000ms after the auditory description of the target, giving listeners different amounts of time to process the expressions before beginning the visual search.

The results showed a different picture for color and pattern, as well as for the timing of the visual presentation: For color discriminability, when it was fully discriminating (Condition A on the left in Figure 2.5), the eye-tracking results showed that Pattern-Color expressions (i.e., overspecifying color) facilitated visual search, indicated by the increased proportions of fixations towards the target compared to the pattern-only description (i.e., MS), which suggested the facilitation effect of informative color OS. When color was only partially discriminating (Condition B), color as the second adjective did not hinder comprehension, as in the results, the proportion of fixations towards the target after hearing a full expression was similar to hearing the pattern-only description. However, the reaction times in both conditions were slower when hearing OS than MS, which, on the other hand,

suggested a hindering effect. Interestingly, if the visual display was presented 1000ms later, the reaction times became faster for OS than MS, and more fixations towards the target were detected in both conditions. This suggested that when the visual display was presented later, the facilitation effect of overspecifying color as the second adjective was further enhanced, probably because the delay provided more time for listeners to incorporate the color information.

By comparison, for pattern discriminability (the right panel in Figure 2.5), the results showed that overspecifying pattern as the second adjective hindered comprehension regardless of discriminability, with a decreased number of fixations towards the target as well as increased reaction times of target identification when guided by OS than by MS. This was also the case when the display was presented later. This systematic comparison between color and size OS showed that the facilitation effect tends to occur for the more salient property (e.g., color) that can help to discriminate the target faster than the alternative properties (e.g., pattern).

To date, the debate over whether OS facilitates or hinders the listener's target identification remains unresolved. Even in studies that concluded a facilitation effect, inconsistencies and varying effect sizes were observed in the experimental results (e.g., Arts et al., 2011; Rehrig et al., 2021). Overall, evidence so far seems to suggest that OS can facilitate comprehension, especially when the property is informative and salient, as it can efficiently help listeners to visually narrow down the number of potential target candidates. But it is also noteworthy that the facilitation effect has only been found under certain visual conditions and for certain visual object properties, for example, with color OS facilitating more effectively than other properties such as size and material (Davies & Katsos, 2013; Ronderos et al., 2024), pattern (Fukumura & Carminati, 2021; Tourtouri et al., 2019), and object status (the *unmelted* candle in Gann & Barr, 2012). The facilitation effect is also more likely to be observed in complex visual scenes, such as real-world environments or 3D virtual spaces, or in demanding tasks that may exceed the listener's processing capacity (see also Sonnenschein, 1982; Sonnenschein, 1984). In simple contrastive scenes (e.g., with only two objects in Engelhardt et al., 2011), on the other hand, OS does not help, but instead, may add confusion for listeners during comprehension.

2.1.2.2 Do Speakers Produce Overspecification with Listener-Oriented Intentions?

Having established in the previous section that informative OS often facilitates target identification for listeners, the next question is whether speakers produce OS with the intention of helping their interlocutors identify the target (i.e., a listener-oriented intention), rather than merely as heuristics to ease their own production. This is a challenging question to address, as research must not only detect the frequent use of OS in certain conditions but also provide evidence that can distinguish this behavior from a purely speaker-oriented intention.

The Incremental Efficiency Hypothesis Paula Rubio-Fernández and her team proposed the Incremental Efficiency Hypothesis, arguing that frequent OS supports the listener-oriented view, as speakers adopt OS to facilitate listeners' efficient visual search (e.g., Long et al., 2021; Rubio-Fernández, 2016, 2019, 2021; Rubio-Fernández et al., 2021). This hypothesis builds on the premise that referential communication should be considered as a collaboration process (Clark & Wilkes-Gibbs, 1986), where speakers and listeners are "co-present", not only by sharing common ground about the visual display they jointly inspect, but also by sharing assumptions and consensuses about the collaboration process as well as the perceptual structure of the visual scene itself (Long et al., 2021; Rubio-Fernández, 2016, 2019, 2021). Among these shared assumptions, two are particularly central to the Incremental Efficiency Hypothesis: the **incremental** nature of the visual search process, and how to achieve **efficient** target identification.

First, the Incremental Efficiency Hypothesis assumes that both speakers and listeners treat visual search as an incremental process. Rather than encoding a referential expression as a complete unit to be matched with the visual scene, speakers produce expressions incrementally, based on the assumption that listeners will process each property in the order it is mentioned. Each property is expected to help listeners progressively narrow down the selection scope until the target can be uniquely identified, which has been supported by evidence for incremental visual search processing in comprehension research (e.g., Eberhard et al., 1995; Rubio-Fernández et al., 2021; Tourtouri et al., 2019). Whether an expression is MS or OS matters less from this incremental, forward-looking perspective, since retrospectively defining the relationship between the expression and the visual scene as MS or OS misses the collaborative and step-by-step nature of the process.

An important piece of evidence for incrementality comes from Experiment 1 in Rubio-Fernández (2016). Using a referential communication paradigm, she compared four-object displays with uniform or multiple colors (monochrome vs. polychrome) across English and Spanish speakers. Unlike in English (e.g., "the *blue* dress"), color modifiers in Spanish are often placed in the post-nominal position (e.g.,"el vestido *azul*" [the dress *blue*]). Her study showed that post-nominal OS in Spanish was significantly less frequent than pre-nominal OS in English for both display types, but even less so in monochrome displays. This outcome has been replicated in Rubio-Fernández (2019) and Rubio-Fernández et al. (2021) by using displays with a greater number of objects and color variations (8-16 objects). The less frequent color OS encoded in the post-nominal structure suggests that when the noun, uttered before color in Spanish, can already identify the target, speakers would assume that the post-nominal color modification may be less helpful for the listener, thus employ it less frequently. In English, by comparison, where color appears pre-nominally, color OS provides more benefit for narrowing down the selection scope of potential referents during visual search. The different frequencies of color OS in the two nominal positions emphasize the importance of incrementality, i.e., the order in which properties are mentioned in referential communication.

The second shared assumption between speakers and listeners is their mutual understanding of how to **efficiently** complete the target identification task. In the Incremental Efficiency Hypothesis, "helpful" or "efficient" does not strictly refer to

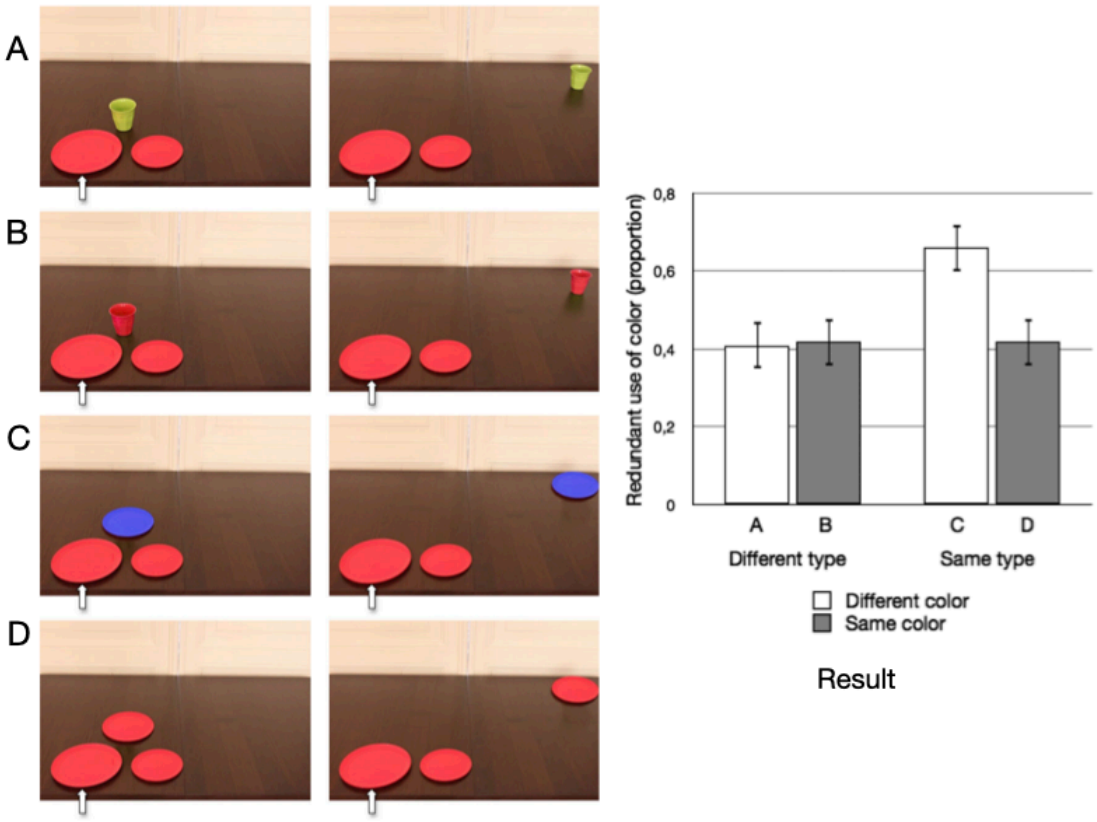


Figure 2.6: Example stimulus and experimental results in Koolen, Krahmer, et al. (2016), Figure 5 (left) and Figure 7A (right). The target is "the big red plate", where "red" is a color OS. Speakers only overspecified more often when the competitor was of the same type as the target but differed in color, i.e., in Condition C. Reproduced with permission from *Cognitive Science*, © 2015 John Wiley & Sons. License number: 6018800497181 (issued April 30, 2025).

the actual speed of target search or whether listeners genuinely benefit from the overspecified information. Instead, it reflects the cooperative intention of speakers and the expectation that listeners will benefit from their expressions based on this shared understanding of efficiency. If speakers consider listeners' demands of efficient visual search, speakers' OS are listener-oriented (Rubio-Fernández, 2016).

Specifically, Rubio-Fernández argues that both interlocutors should mutually agree that the visual perception of the property *color* is efficient, as color is visually salient, compared with most of the other visual properties adopted in a referential display (e.g., size, pattern, material). Speakers should assume that listeners can efficiently incorporate color information in their target search, serving as an effective visual heuristic for both production and comprehension (Long et al., 2021). The efficient visual nature of color can explain why color OS is more frequent in visually complex scenes, such as polychrome displays and those with a greater number of objects and color variation, as well as when describing a color pop-out target (using stimuli similar to Figure 2.3 in Long et al., 2020, 2021; Rubio-Fernández, 2016, 2019, 2021).

Although the Incremental Efficiency Hypothesis proposed by Rubio-Fernández is appealing at the theoretical level, the experimental conditions and results presented in her studies fall short of fully justifying her accounts, as they do not always clearly distinguish themselves from the speaker-oriented view. Based on her stimuli and conditions (similar to Figure 2.3), the speaker-oriented view would also predict a similar, high frequency of color OS, especially in visually complex displays with color variations, where color can often directly identify the target in their experimental conditions.

Furthermore, not all empirical evidence aligns with the predictions of the Incremental Efficiency Hypothesis: the findings from the study by Koolen, Krahmer, et al. (2016) seem at odds with the Incremental Efficiency Hypothesis (a similar result was found in Koolen, 2019, but the stimuli were presented in 3D virtual reality). In Experiment 2 of this study, they manipulated whether a competitor differed in color or type from the target, and also varied the physical distance between competing objects (Figure 2.6). The results showed that, regardless of distance, speakers overspecified color more frequently only when the competitor was the same type but a different color than the target (i.e., more color OS in Condition C in the results in Figure 2.6). If speakers truly planned their utterances incrementally, as the Incremental Efficiency Hypothesis predicts, color OS should have been equally frequent in Conditions A and C, because color OS should facilitate target identification regardless of the competitor's type.

However, the results showed that speakers also considered object type (the noun) when deciding whether to overspecify, the property encoded after the color adjective.[1] This suggests that incrementality alone does not fully account for the observed patterns of OS, and that the noun can also retrospectively decide whether the pre-nominal color OS was necessary. It is also important to note that the results of this study may not fully align with predictions made under the speaker-oriented view either. Condition A, which had the highest scene variation (varying in both color and object type), did not yield the highest frequency of color OS as the heuristic account might expect (e.g., Koolen et al., 2013). The result of this study also suggests that speakers' encoding strategies may differ when the visual stimuli are simple, with only three objects, compared to scenes that are more visually complex, probably because speakers may have more cognitive capacity to plan their utterances when dealing with simpler visual inputs.

While the findings in Koolen, Krahmer, et al. (2016) pose challenges for the predictions of the Incremental Efficiency Hypothesis, subsequent studies have sought to support this hypothesis through more carefully controlled and refined experimental designs. For example, S. A. Wu & Gibson (2021) ruled out language-specific production tendencies as an explanation for OS patterns between English and Spanish speakers, e.g., English speakers may tend to be verbose in general. They compared OS usage between color modifiers, which are pre-nominal in English and post-nominal in Spanish, and cardinality modifiers (e.g., "*three* circles"), which are encoded in the pre-nominal position in both languages. If the reduced post-nominal OS in Spanish were due to cultural or linguistic backgrounds, i.e., Spanish speakers overspecify less in general, they would also show lower OS

[1]The experiment was conducted in Dutch, where color is encoded pre-nominally.

rates for cardinality. However, the experimental results showed that they did not: both groups had equal OS rates for cardinality modifiers, while Spanish speakers produced less color OS in the post-nominal position. This further supports the Incremental Efficiency Hypothesis and excludes the confounding effect of language-specific preferences for OS.

The Role of Informativity on Overspecification Beyond the Incremental Efficiency Hypothesis, recent studies have begun to examine whether the concept of informativity, closely tied to communication efficiency, can further distinguish between the speaker-oriented and listener-oriented accounts of OS. If evidence shows that speakers consider informativity when using OS, it would already suggest that OS is indeed listener-oriented, as this consideration implies that speakers are taking the listener's need of efficient visual search into account.

For instance, Mangold & Pobel (1988) varied the degree of a property's discriminability relative to the target (as in fully-, partially-, or non-discriminative, representing a qualitative form of informativity). They found that speakers significantly overspecified partially discriminative properties more frequently than non-discriminative ones, particularly when the partially discriminative property was perceptually more salient (e.g., color). Davies & Kreysa (2017) paired speakers' eye movements with their utterances and found that, although not mandatory, informative descriptions[2] are often enhanced by speakers looking at the contrasting competitor before speaking (e.g., looking at "the small desk" before describing "the big desk"). These eye movement patterns suggest that speakers do not simply fixate on the target for the referential task; rather, they consider informativity by looking at the contrastive competitor to meet the "pragmatic expectations of listeners". Such utterances also resulted in longer speech onset times, suggesting that speakers exert extra effort to tailor their descriptions to the listener's needs.

Tourtouri et al. (2019) investigated color and pattern OS by manipulating the RER of these properties in referential communication tasks. Across six conditions (Figure 2.7), while one of the properties (e.g., "blue" or "striped") was sufficient to minimally specify the target (e.g., "the blue/striped ball") across conditions, the unnecessary property varied in its informativity based on its RER. Crucially, in Conditions b) and e) in Figure 2.7, the unnecessary adjective reduced even more entropy than the necessary adjective (narrowing down the referents to 2 objects instead of 4). If informativity can influence OS encoding, one would expect not only a high frequency of OS across conditions, but also ideally, that the OS rate would be proportional to the RER of the unnecessary property: i.e., higher OS rates would be found when the unnecessary adjective carried a higher RER compared to a lower RER.

The results revealed three distinct patterns in speakers' behavior. One group consistently overspecified all properties across every condition, regardless of informativity. A second group frequently overspecified color across conditions but did so less often for pattern OS. The third group showed the pattern expected

[2]Noting that given the design, Davies & Kreysa (2017) primarily compared MS and underspecification. Informative descriptions here meant the MS of the target. There was no room for OS given the stimuli. Little is known about speakers' eye movement when producing OS.

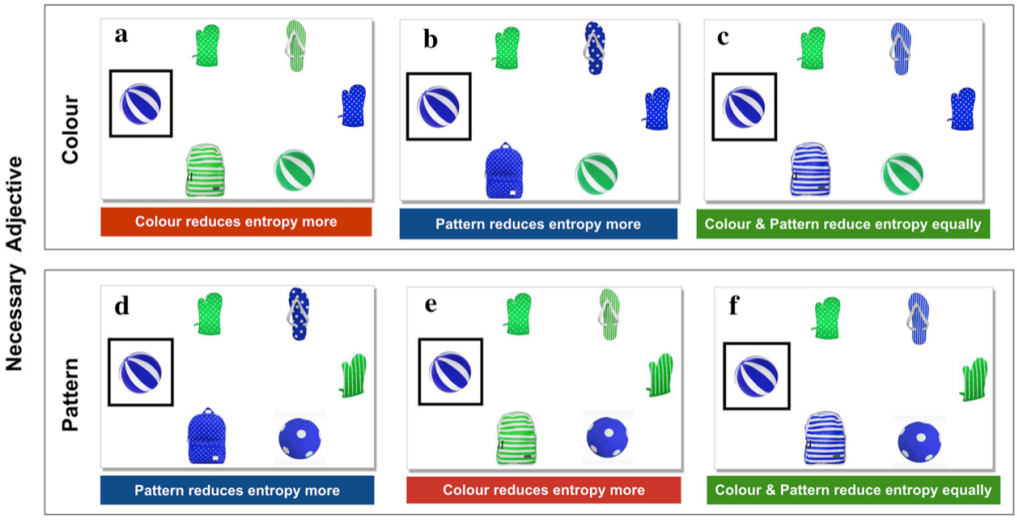


Figure 2.7: Example stimulus in Tourtouri et al. (2019), Figure 8. Only using the necessary adjective (color or pattern) can already minimally specify the target. The unnecessary property differed in its informativity based on RER. Reproduced from the original figure under the terms of CC BY 4.0.

by the listener-oriented view, where OS rates were proportional to the RER of the unnecessary property, although color OS tended to be slightly more frequent overall.

This was the first study showing that, at least for a subset of speakers, they overspecify *strategically* according to RER of each property to facilitate target search for listeners, by overspecifying more frequently the informative, high RER property, which is more useful for visual search. At the same time, the observed group differences suggest that both the speaker-oriented and listener-oriented views might be valid, revealing that speakers may employ OS either as a heuristic or as a listener-oriented strategy. If more studies can replicate these results across other visual properties and different RER contrasts, future research may start to shift from debating which view is more valid into investigating which types of speakers prefer which strategy and under what circumstances.

To summarize the listener-oriented view, empirical evidence from perspectives of both interlocutors indicate that speakers produce OS to assist listeners in their visual search for targets, with informativity playing a crucial role in this collaborative process. From listeners' perspective, informative OS enables them to narrow down their visual search for target referents more efficiently at an earlier stage. From speakers' perspective, they design their utterances to address listeners' needs for incremental and efficient visual search, resulting in more frequent OS for the informative properties. It is also important to note that informative OS, and the resulting facilitation effect for listeners, is more likely to occur in visually complex scenes and favors properties that can more effectively distinguish the target referent, such as color.

2.1.3 Discussion: Moving Away from Property Selection

The previous two sections reviewed studies focusing on property selection, particularly OS, as a case study to examine the speaker-oriented and listener-oriented views, as well as the role of informativity in referential communication tasks. Comparing the two views, the key difference lies in the communication purpose behind the use of OS by speakers: it can either stem from a heuristic drive for fast speaking or represent a strategic act aimed at facilitating listeners' target identification. Informativity plays a crucial role in clarifying this divergence. OS produced with a heuristic motive should not be influenced by informativity, as speakers may not have thoroughly inspected a complex visual scene before speaking. In contrast, OS generated with a listener-oriented intention should take the informativity of property words into account, since informative property words can facilitate listeners to identify the target more efficiently.

2.1.3.1 Limitations of Studying Overspecification

While OS is a classic phenomenon in referential communication, it has its limitations as the case study in verifying the speaker-oriented versus the listener-oriented views: One of the biggest challenges is to use the OS rate as the measure of interest to tease apart the two diverging views. While both the speaker-oriented and listener-oriented accounts adopted OS as evidence to support their views, the results and the interpretations have been heavily mixed and confounded. Sometimes, the same results were adopted to account for different hypotheses: e.g., higher color OS rate in visual complex scenes is both predicted by the speaker-oriented view (e.g., Koolen et al., 2013), and by the listener-oriented view (e.g., Rubio-Fernández, 2019), although the reasoning behind was drastically different.

To truly tease apart the two views, in addition to systematically examining the role of informativity (e.g., Tourtouri et al., 2019), previous research has also attempted to tackle this issue by, e.g., adding the task difficulties (e.g., Gann & Barr, 2012; Vogels et al., 2020), and pairing OS with other measures, such as speech onsets. However, the results of these studies are largely mixed and indirect for interpretation.

The motivation for increasing task difficulty is to examine whether speakers can maintain a listener-oriented approach under high task loads, either on speakers or listeners. When task difficulty increases for speakers, a heuristic speaker may struggle to keep track of the listener's needs and revert to easy production (e.g., Bock, 1982; MacDonald, 2013), whereas a listener-oriented speaker may exert additional effort to assist listeners by e.g., inspecting the visual scene more carefully and strategically designing their utterances; When task difficulty increases for listeners, a heuristic speaker may neglect to consider listeners' challenges during reference production, while a listener-oriented speaker should recognize the increased difficulty for listeners and tailor their descriptions to ease listeners' task.

One way to increase speakers' task difficulty is to make the target harder to describe. Recall the study by Gann & Barr (2012), they included not only conventional targets (e.g., the unmelted candle) but also objects with unconventional shapes (Figure 2.8). Since speakers lacked prior experience describing these

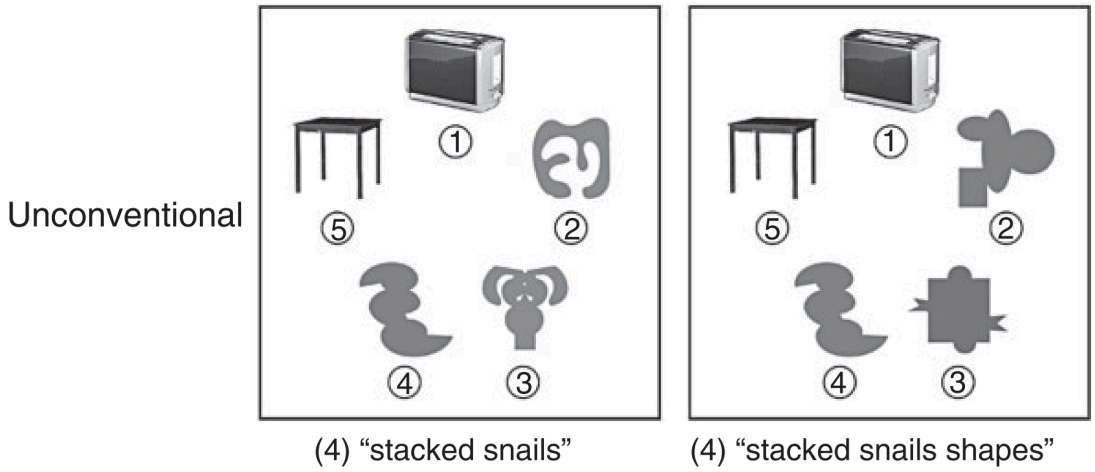


Figure 2.8: Example unconventional targets in Gann & Barr (2012), Figure 1. The display on the left is the stimulus for the training phase, and the display on the right is for the testing phase. Reprinted from *Speaking from experience: Audience design as expert performance*, by T. M. Gann and D. J. Barr, originally published in *Language, Cognition and Neuroscience*, 29(6), 744–760. © 2012, reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group, https://www.tandfonline.com

unconventional objects, they had to create naive descriptions during the training phase. Alongside the manipulation of old versus new addressees, the unconventional targets presented in the testing phase may either be the same as those from the training phase or completely new. The key question is whether speakers would directly reuse the descriptions created during the training phase in the testing phase, supporting the speaker-oriented view, or adjust their descriptions based on the addressees' status (old or new), supporting a listener-oriented view.

Interestingly, the results indicated that participants tended to use more words when describing a *new* referent to *familiar* addressees, with longer speech onsets suggesting greater planning effort for these utterances. However, the word length for new addressees showed no significant difference between old and new referents. If longer descriptions can be viewed as a form of OS, this finding suggests that, especially for unconventional new targets, speakers prefer to provide more information to help the familiar addressees distinguish the new targets, thereby supporting the listener-oriented view. However, the unchanged word length when describing the old referents to the new addressees seems to suggest that speakers are still heavily influenced by their prior experiences from the training phase, which can be interpreted as a heuristic approach.

To increase listeners' task difficulty, Vogels et al. (2020) designed a driving scenario that incorporated a referential communication task. In this setup, listeners acted as drivers, guided by speakers who were passengers in a virtual driving environment. Alongside the target identification task, listeners faced varying levels of driving difficulty: either easy (driving straight) or difficult (turning wheels based on complex cues). The two parties swapped roles as listener-drivers and speaker-passengers in the middle of the experiment. If speakers are indeed listener-oriented, one would expect them to produce more OS when the listener-driver is

engaged in a challenging driving task. However, the results did not support this prediction; speakers' OS patterns remained consistent regardless of task difficulty. Interestingly, participants who first experienced the difficult driving task as listener-drivers subsequently adopted more OS when they took on the role of speakers. This suggests that OS is more likely to be adopted when strong cues, such as firsthand experience in the listener role, are presented initially.

By adding task difficulties, both Gann & Barr (2012) and Vogels et al. (2020) only partially supported the listener-oriented view: it seems that evidence of OS was mixed in supporting the two views, while speakers still show consistent preference of using OS as a heuristic, they also consider the listener's perspective, but only under specific conditions and tasks.

Another reasonable measure to further tease apart the two diverging views is to pair speakers' referential descriptions with their **speech onset latency**, which captures the time speakers take to inspect the visual scene and plan their utterances before speaking (e.g., Brown-Schmidt & Konopka, 2011). Speech onset may be correlated with the effort speakers put into these processes. It would be reasonable to speculate that an OS produced with the listener-oriented intention would result in a longer speech onset compared to the same OS produced through a heuristic strategy (also hypothesized by e.g., Gann & Barr, 2012; Vogels et al., 2020).

However, not many previous studies have recorded speech onset data, and the results among those that did are heavily mixed or confounded. For example, Gann & Barr (2012) detected longer speech onsets for longer descriptions, specifically when speakers described new referents to familiar addressees (discussed above). However, speakers did not show longer speech onsets when designing their descriptions for the new addressees, suggesting no extra effort was made to tailor the description for listeners, who encountered the unconventional targets for the first time. Vogels et al. (2020) found shorter speech onsets for OS compared to MS, implying that producing OS may actually be less effortful than MS and therefore adopted as a heuristic strategy (see also Belke & Meyer, 2002 discussed in Section 2.1.1). Potential support for the listener-oriented view comes from findings by Long et al. (2021), where color OS was associated with longer speech onsets in polychrome scenes compared to monochrome scenes. However, since this manipulation was confounded with visual complexity, it remains unclear whether the longer onsets reflected a listener-oriented strategy or simply the cognitive cost of processing a visually complex display.

In sum, it appears that, aside from studies investigating the effect of informativity (e.g., Tourtouri et al., 2019), neither speech onsets nor task difficulty manipulations have been sufficient to clearly distinguish between the speaker-oriented and listener-oriented views for OS. Speech onsets may reflect differences in the production effort of OS, but they do not reliably indicate a listener-oriented intention. It remains possible that both views capture valid aspects of referential production, with OS adopted for different purposes depending on speakers' cognitive capacity, task demands, and prior experience.

2.1.3.2 Key Takeaways

The discussion so far shows that OS can be accounted for by both the speaker-oriented and listener-oriented views, as speakers may adopt the same OS strategy with either underlying intention. This project proposes moving beyond the focus on OS and property selection to explore **property ordering**, or **linearization**, of reference production, in order to better distinguish between the two diverging views.

The motivation for studying property ordering arises from the well-established understanding that different word order alternatives for meaning-equivalent expressions typically involve a *preferred order*, which is most commonly used in everyday language, and other *less preferred orders*, which are less frequent and often more effortful to produce. Examples of this include SOV versus SVO structures (e.g., Branigan et al., 2008) and direct-object versus prepositional-object alternations (e.g., Branigan et al., 2000). In referential communication, this applies to **pre-nominal** and **post-nominal structures**. Producing the less preferred order may require more cognitive effort (e.g., Brown-Schmidt & Tanenhaus, 2006), and speakers may adopt it only for specific reasons (e.g., priming, Cleland & Pickering, 2003; to be informative, Fukumura, 2018). The following Section 2.2 will delve deeper into this issue of property ordering. Before that, several key insights from OS research discussed earlier will be applied to the study of property ordering:

First, the current project broadly aligns with the Incremental Efficiency Hypothesis (Rubio-Fernández, 2016), which posits that speakers assume listeners' visual search is incremental, building on the order in which the property words are presented. Thus, speakers are likely to order the properties in a sequence that can efficiently facilitate incremental target identification. From this perspective, property ordering plays a more central role than property selection, as incrementality is achieved through ordering and linearization.

Second, the current project assumes that an informative property that significantly narrows down the selection scope of referents is more efficient for the visual search of the target. Therefore, this project will follow Tourtouri et al. (2019) and quantify informativity as RER to link the selection scope with the information-theoretic notion of entropy reduction. The project hypothesizes that a speaker-oriented strategy would rely on the preferred order regardless of informativity, while a listener-oriented strategy would vary word order based on informativity. This listener-oriented approach would order properties in a sequence that can facilitate visual search, even when this means that speakers need to produce the more effortful, less preferred order.

Third, building on findings from Tourtouri et al. (2019), the project expects that both speaker-oriented and listener-oriented strategies exist among speakers. However, it remains unclear whether this distribution is fixed in the population or changes under varying task demands (e.g., speakers may seek heuristic strategies when the task becomes more difficult). This project also aims to explore whether and when speakers would switch between heuristic and listener-oriented strategies in different communication environments.

2.2 Property Ordering in Reference Production

The previous section discussed property selection under the debate of the speaker- and listener-oriented views. This section, moving away from property selection, explores what we have known so far about property ordering in referential production. It will mainly review two research topics about property ordering: one regarding *adjective ordering*, where multiple properties are encoded pre-nominally. Theories about adjective ordering have been extensively studied in multiple fields to account for why adjectives are ordered the way they are and why certain ordering is preferred over others. Many studies have also touched upon the role of informativity in deciding the orders.

The second topic regards studies that adopted the referential communication task to elicit *post-nominal modifications*. Very few studies so far have systematically studied the ordering of post-nominal modifiers. These studies will be introduced, not only because the current project compares pre- and post-nominal structures, but more importantly, because the post-nominal structure, compared with the pre-nominal structure, involves an ordering alternation at the **syntactic level**, which is assumed to be qualitatively different from alternating adjective orders that only happens in the pre-nominal position. By qualitative, I refer to the potentially increased difficulty in planning to produce this structure (e.g., Brown-Schmidt & Konopka, 2008), and the fact that the ordering of post-nominal modifiers appears to be less rigid compared to pre-nominal ones (e.g., Trainin & Shetreet, 2021).

This section will discuss these two topics from three perspectives: first, the most influential factor of word order is the conventional constraints of language. With a strict convention, the order is often fixed rather than flexible. Section 2.2.1 will introduce several principles that constrain the order of adjectives, corresponding to the common properties adopted for the objects used in referential displays. Section 2.2.2, as the central perspective for this section, examined the psycholinguistic accounts for property ordering. Although adjective ordering is constrained by grammatical and semantic conventions, psycholinguistic studies, however, have shown that ordering can be flexibly alternated by certain psycholinguistic factors. Informativity plays an essential role here. Section 2.2.3 will explore information-theoretic accounts related to RER and discuss their predictions for property ordering. Taking these perspectives together, the goal of this section is to show that although property ordering is, by convention, fixed and rigid, certain factors, such as informativity, can lead to alternative ordering, for e.g., efficient communication.

2.2.1 Conventional Constraints

The study of adjective ordering can be dated back to Sweet (1898). In this English grammar book, Sweet proposed two central principles that determine adjective ordering based on the semantic meaning of adjectives in relation to the noun. The first principle is called *closeness to the noun in meaning*: "When a noun has more than one modifier, the general principle that the one most closely connected to it in meaning comes next to it" (Sweet, 1898, p. 8). For instance, in the noun phrase *the three wise men*, *wise* is semantically closer to describe the quality of *a man*,

while the numeric modifier *three* is less semantically relevant (probably because it can also modify inanimate objects, e.g., *three* boxes, example from Scontras, 2023). The second principle is *definiteness of denotation*, distinguishing whether adjectives modify the *inherent* qualities of the noun (see also Byrne, 1979; Whorf, 1945). A more inherent quality should appear closer to the noun, such as the color and material of objects. Other qualities, such as size and quantity, appear further away from the noun. Martin (1969) further developed the two principles by introducing concepts such as the *associate strength* between adjective and noun meanings, and the *absoluteness* of the quality that the adjective attributes to the noun. One of the latest theories by Scontras et al. (2017) suggests that the *subjectivity* of properties denoted by adjectives can unify various ordering principles. According to this view, less subjective adjectives are positioned closer to the noun; for example, English speakers tend to use "the big blue box" rather than "the blue big box", because deciding the size of an object is a more subjective judgment than color.

These principles suggested that depending on the quality an adjective denotes, adjectives with the more absolute, inherent, definite, and objective qualities will appear closer to the noun. Greenberg (1963) even hypothesized that these principles are universal and should be applied cross-linguistically (see also Scontras et al., 2017). For languages that encode modifiers in the post-nominal positions, e.g., in Spanish, the principles should still be applicable, by reversing the order so that modifiers depicting the more inherent qualities are still placed closer to the noun, known as *mirror-image ordering*, e.g., "the plate that's red and big" in a post-nominal based language (e.g., Cinque, 1994; Scontras et al., 2017).

However, it is noteworthy that these principles and the assumption of universality are largely concluded based on Germanic and Romance languages. Whether the principles can be applied in other languages is still debatable: a counterexample has been reported by Alghazo & Jarrah (2023), who detected no strong adjective ordering preferences in Jordanian Arabic. Also, the ordering of post-nominal modifiers does not always follow the mirror-image ordering, which has been reported in Hebrew (Trainin & Shetreet, 2021; but also see supporting evidence in e.g., Modern Standard Arabic in Atiega et al., 2022).

Even among Germanic and Romance languages, the strength of these principles may not be as high as those found in English (see an analysis for German adjectives in Kotowski & Härtl, 2019). For example, comparing German and English, while a color modifier is significantly preferred to be placed in the pre-nominal position than in the post-nominal position in English (e.g., "the red car" vs. "the car that is red"*, Lester & Beason, 2013). In German by contrast, it is also natural to place the color modifier in an relative clause in the post-nominal position (e.g., "das rote Auto" [the red car] vs. "das Auto, das rot ist" [the car, that is red], Durrell, 2011). The level of flexibility of ordering preferences may still render large cross-linguistic differences (Sproat & Shih, 1991).

Regardless of the detailed cross-linguistic variations, the conventional constraints about adjective ordering highlight the innate semantic relation between adjectives and head nouns, which would predict that at least pre-nominally, adjective ordering should be largely rigid and fixed within a language based on semantic hierarchies (Scott, 2002), because the denoted qualities of adjectives already determine the

distance to the noun (e.g., a size adjective would appear farther than color to the noun).[3] Previous studies about OS, as introduced above, rarely reported the ordering dimension of speakers' production results, which would, unless otherwise specified, indicate that speakers followed the preferred adjective ordering, such as "the big red plate" in Gatt et al. (2017), where color appears closer to the noun than size.

2.2.2 Psycholinguistic Accounts

Many psycholinguistic studies have questioned the rigidity of adjective ordering and explored alternative factors that influence the preferred order, as well as those that can regularly invert it. This section examines the roles of incrementality, priming, and, most importantly, informativity in shaping the ordering of referential expressions. Perspective-taking may also play a role, as suggested by indirect evidence from broader communication tasks.

This section will also review previous psycholinguistic research on post-nominal modifications in referential communication tasks, as the current thesis project, under certain conditions, predicts a shift from the preferred pre-nominal modification to the less favored post-nominal order. Evaluating the feasibility and challenges of such an alternation is important for estimating the potential effect of informativity and its expected effect size.

2.2.2.1 Incrementality

A very important psycholinguistic force deciding ordering is **incrementality**. Incrementality here can refer to the process of both comprehension and production. For incremental comprehension, similar to what has been introduced before in Section 2.1.2.1, listeners' target visual search is incremental, time-locked to the word order (e.g., Eberhard et al., 1995). A collaborative, listener-oriented speaker would be aware of this incremental process of the listener and thus plan their expressions accordingly.

Studies adopting a visual world paradigm have confirmed this incremental process in **comprehension**: Rubio-Fernández et al. (2021) and Rubio-Fernández & Jara-Ettinger (2020) examined the ordering effects on visual search among participants with different language backgrounds: English, Hindi, and Hungarian as the pre-nominal users, and Spanish, Catalan, and Wolof as post-nominal users. Both studies observed that listeners indeed visually searched for targets by exploiting the incremental word order, that is, pre-nominal users first looked more at the pre-nominal adjective candidates (e.g., two *short* objects, or two *blue* object), while post-nominal users compared objects directly (i.e., the noun).

[3]Here is not to claim that the inverted ordering is strictly prohibited. The less preferred, unconventional orders can also be acceptable to readers, depending on the specific rating task (Leivada, 2022; Leivada & Westergaard, 2019; but see Engelhardt et al., 2008 reporting low acceptability). The focus here is that the likelihood of producing the unconventional adjective order would be scarce compared to those that conform to the conventions (e.g., Greenberg, 1963).

Informativity also plays an important role in this incremental comprehension process. By alternating the adjective order of color and pattern, Fukumura & Carminati (2021) detected a greater amount of fixations allocated towards the target as well as faster reaction times for target identification, when the first adjective is more discriminating, i.e., more informative, especially when the first informative adjective is color.

For **production**, it is widely agreed that language production is incremental: speakers often begin with the more available and accessible elements to initiate speech early, while continuing to plan the remaining part of an utterance during articulation (e.g., Bock, 1982; Brown-Schmidt & Konopka, 2008; F. Ferreira & Swets, 2002; MacDonald, 2013; Martin, 1969; Pechmann, 1989). In the domain of adjective ordering, a more available or accessible adjective is the one that is perceptually more salient (e.g., color, Bever, 2013; Bramão et al., 2011; Clarke et al., 2013; Long et al., 2021), or linguistically more frequent (e.g., "wooden" vs. "crocheted" in Scontras et al., 2017; see also Martin, 1969).

Pechmann (1994) and Pechmann & Zerbst (1990) registered more frequent use of the less preferred adjective ordering of color-size when the visual complexity of a referential display was higher. They reasoned that it was because, under time pressure and higher task demands, the more salient color adjective should be accessed earlier than size in incremental production. Belke (2006) further introduced another form of task difficulty by restricting one of the subject groups to only produce MS for targets, while the other group can describe the target freely, allowing OS for certain conditions. Only producing MS was assumed to be a more difficult task because speakers were required to compare the target with every other object in a visual display (Belke & Meyer, 2002). More color-size ordered expressions were detected under the MS production task, presumably because visual comparison of color and size properties became more refined, resulting in a greater emphasis on color. However, they did not detect increased use of color-size order under time pressure that was predicted by Pechmann (1994).

To date, the extent to which incrementality plays a role in adjective ordering is still debated. The unsolved questions include, for example, although being successfully inverted under certain conditions by Pechmann (1994) and Belke (2006), the overall preferred order of size-color (at least in English) is still in stark contrast to the incremental account, where color should be more salient and accessible earlier than size. Another unresolved question concerns the interaction between incrementality and the ordering principles for post-nominal modifiers. If, as suggested by incremental production theories, more accessible modifiers are generally preferred to be placed earlier in utterances, this principle should theoretically apply to both pre-nominal and post-nominal adjective orders, predicting parallel orders across both structures. However, in many languages, post-nominal modifiers tend to follow the mirror-image ordering relative to pre-nominal structures, as conventionalized by grammar and supported by empirical findings (see a discussion by Scontras, 2023). These topics warrant further empirical investigation.

2.2.2.2 Priming

The second influential factor on adjective ordering is **priming**. Priming refers to speakers' tendency to repeat or reuse linguistic units or structures that they have previously been exposed to (e.g., Bock, 1986; Hartsuiker, 1999). Unlike pure lexical or syntactic priming, the priming effect in referential communication highlights a tendency for interlocutors to gradually align with the behavior of each other through interaction (Pickering & Garrod, 2004). It has been shown that speakers tend to align with their partners' referential behavior in producing MS or OS, above and beyond the simple explanation of syntactic priming (Loy & Smith, 2021). They also tend to select the primed property, even when it was the less preferred property (e.g., direction, as in "a desk facing backward"), among other options (e.g., color, as in "a red sofa" in Gatt et al., 2011).

Priming can also alter the order of adjectives: the frequency of less preferred orders increases when a conversation partner have used the same order in previous trials (e.g., speakers produce more expressions such as *"the red big sofa"* when primed in Goudbeek & Krahmer, 2012). Priming of modifier ordering can also occur cross-linguistically for bilingual speakers, as shown by Adamou et al. (2021), who found that priming the same adjective order from Romanian to Romani increased use of the less preferred order. It can also influence ordering at the syntactic level, such as increasing the use of post-nominal modifications like *"the square that is red"* (Cleland & Pickering, 2003). Speakers may also engage in self-priming, reusing syntactic structures they have previously adopted (Jacobs et al., 2019). Self-priming is also relevant to the consistency effect reported by Tarenskeen et al. (2015), where speakers were more likely to overspecify size if they had previously produced color OS within an experimental session.

Many studies have simultaneously investigated the effects of priming and informativity within the referential communication environments, so it is necessary to introduce the third influential factor on property ordering upfront, that is, the **informativity**, or the discriminability of a property word.

2.2.2.3 Informativity

The theoretic motivation of informativity is similar to what has been introduced in research of OS (see Section 2.1.2): if speakers consider that listeners' visual search is an incremental process, speakers would prefer to place the more informative adjective in an *earlier* position to efficiently facilitate target identification. Similar to the notion of RER, informativity can refer to the discriminability of a property, i.e., the extent to which a property helps to discriminate the target among the rest objects.

Fukumura (2018) specifically examined the role of discriminability in determining adjective order for color and pattern within a referential communication task (e.g., "a green dotted bowtie" as the preferred order). She also explored how discriminability interacts with other factors, such as priming and availability during incremental production. The study included three conditions, illustrated in Figure 2.9. In Conditions A and B, the relevant property provided greater discriminability for identifying the target, while in Condition C, both properties were equally discriminating.

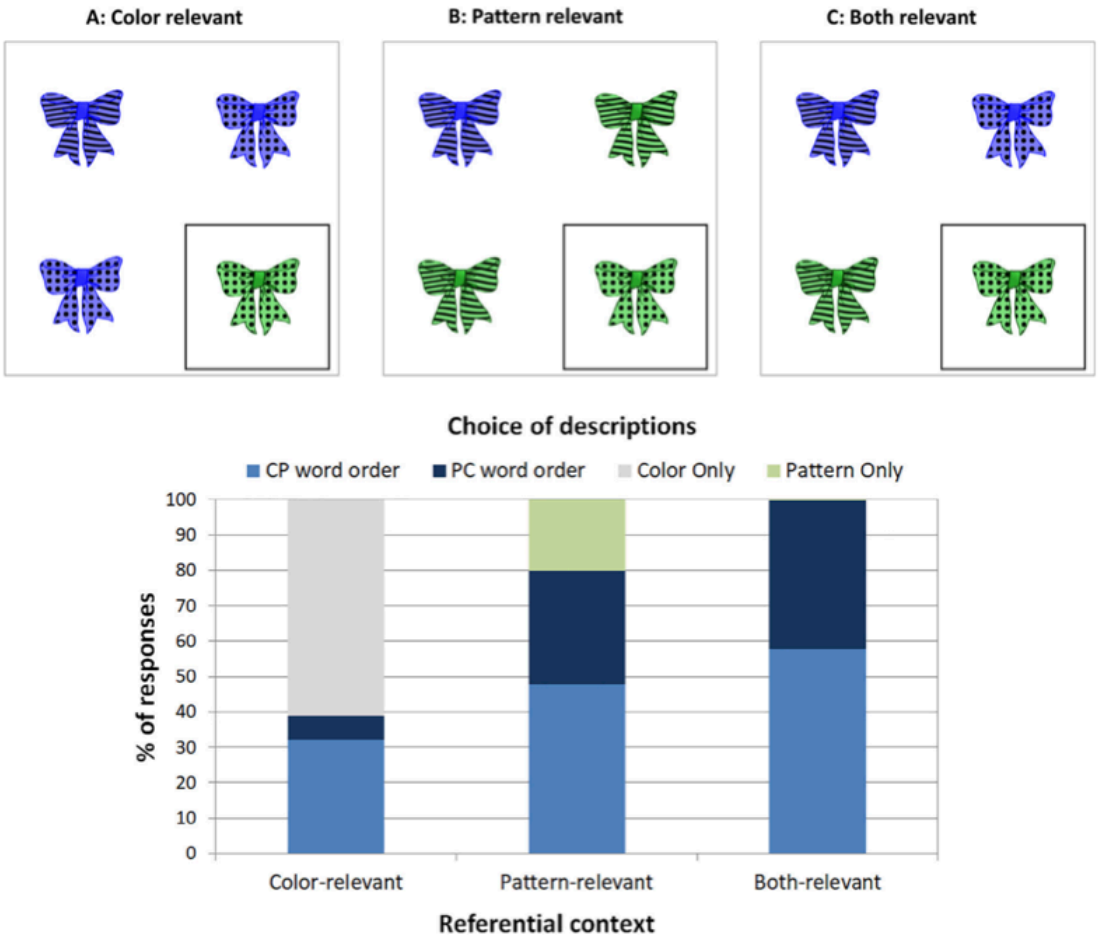


Figure 2.9: Example Stimuli and results of Experiment 1 in Fukumura (2018), Figure 1 (top) and Figure 2 (bottom). The relevant property is more informative and discriminating. "CP" and "PC" in the bar plot refer to adjective orders: Color–Pattern and Pattern–Color, respectively. Reproduced with permission from *Journal of Memory and Language*, © 2018 Elsevier. License number: 6018810135677 (issued April 30, 2025).

In Experiment 1, participants, serving as speakers, described displays across the three conditions. Experiments 2 and 3 introduced priming by pairing participants with a confederate who described targets from Condition C, where both adjective orders (color-pattern and pattern-color) were equally adopted, thereby equally priming the participants. The participants, after a confederate's turn, described targets in Conditions A or B that varied in discriminability. Experiment 3 added a coordination structure in confederate's descriptions, by placing "and" between adjectives (e.g., "a green *and* dotted bowtie").

The results across the three experiments consistently confirmed a significant effect of discriminability on adjective ordering, where the more informative property was more often placed first, especially evidenced by the more frequent use of the less preferred, pattern-color order when pattern was more discriminating in Condition B. This effect was even more robust in Experiment 3, when adjectives were linked

in the coordination structure with the use of "and". Results in Experiments 2 and 3 also suggested a priming effect, where participants frequently repeated the same adjective order adopted by the confederate in the previous trial. There was no interaction between the effect of informativity and priming.

The results of the three experiments also provide evidence for the effect of availability, where color OS, which is possible for Condition B, and the color-pattern order remained frequent overall, because color is more salient and should be available earlier than the pattern adjective, thus is more likely to be uttered first or overspecified in general. Fukumura (2018) concluded these results by introducing the PASS model (Parallel-Activation-for-Selection-and-Sequencing), proposing that both informativity and availability jointly influence property selection and ordering (see more discussion in Chapter 6.2.2.2).

Haywood et al. (2003) investigated the effects of priming and the addressee's role on modifier ordering in a referential communication task. While they did not directly address informativity as strictly defined in the current project, their task is relevant: participants were asked to identify a target card with a specific color and pattern from a deck organized by different priorities. The deck was sorted either by pattern first and then by color (Pattern Box) or by color first and then by pattern (Color Box). Participants collaborated with confederates who were either helpful — using an order that aligned with the box organization (e.g., "find the stripy orange square" for the Pattern Box) — or unhelpful, consistently violating the order of the sorting sequence (e.g., "the orange stripy square" for the Pattern Box). This setup also introduced the effect of priming on adjective orders. In Experiment 2, only the helpful confederate was present, focusing on alternations in color-shape order through pre- and post-nominal modifications (e.g., "the red square" vs. "the square that is red"), using similar manipulations.

The box's sorting sequence could be interpreted as a goal-oriented form of informativity, where the top-level sorting property can facilitate target searching. The hypothesis was that speakers, if considering the needs of listeners, would use adjective orders aligning with the box's sorting sequence more frequently. Results indicated a clear priming effect in both experiments, but showed mixed results based on the effect of sorting sequence. In Experiment 2, with only the helpful confederate, participants tailored descriptions to their partner's needs, prioritizing shape when describing targets in a Shape Box by using post-nominal modifications more often. However, in Experiment 1, participants showed no ordering effect based on sorting sequence, instead, they favored the preferred color-pattern order regardless of the box organization.

Notably, the confederate's helpfulness influenced participants' behavior: they were more likely to modulate the order based on sorting priority when their partner was helpful, but produced less color-pattern order to select from a Color Box when talking with an unhelpful confederate. Though the effects were not robustly consistent, Haywood et al. (2003) identified significant influences of priming, confederate helpfulness, and sorting priorities on modifier ordering, beyond simple syntactic considerations, highlighting how the top-level property can help target identification.

Both studies by Fukumura (2018) and Haywood et al. (2003) suggested that informativity and priming are influential factors in deciding adjective ordering,

especially when communicating with a collaborative partner. Speakers tend to prioritize the informative property in adjective ordering to facilitate target identification, and at the same time, they seek structural alignment with the partner for interactive and collaborative communication.

2.2.2.4 Perspective Taking

Another psycholinguistic factor that may influence property ordering is **perspective taking**, which has been repeatedly shown to shape speakers' utterances. Although the studies reviewed below do not directly address property ordering, they offer insights into how perspective-taking affects referential production more broadly.

In previous research, perspective-taking has been manipulated by introducing substantial variation in how the addressee is represented within the experimental setup. These manipulations require speakers to adjust their own perspective to align with that of the listener to varying degrees: from simply being aware of basic partner characteristics, such as whether the partner is old or new (e.g., Galati & Brennan, 2010; Gann & Barr, 2012), absent or present (e.g., Koolen et al., 2011; Lockridge & Brennan, 2002), a stranger or a friend (Fussell & Krauss, 1989), or collaborative versus unhelpful (e.g., Haywood et al., 2003; Koulouri & Lauria, 2009), to more demanding experimental tasks where participants alternate between the roles of speaker and listener (e.g., Sikos et al., 2021a; Vogels et al., 2020). Other common manipulations include whether the speaker and the addressee share visual access to referential displays or whether the speaker holds *privileged knowledge* (introduced below, see also e.g., Brown-Schmidt, 2009; Damen et al., 2019; Heller et al., 2012; Mozuraitis et al., 2018; Wardlow Lane & Ferreira, 2008); whether communication occurs with a human or a computer (e.g., Peña et al., 2023); and whether the task goal is speaker- or listener-oriented (e.g., *informing* versus *requesting* in Yoon et al., 2012, see Chapter 5). These studies have shown that speakers often adjust their utterances based on the characteristics and needs of their interlocutor, for instance, omitting information when addressing a familiar addressee (Galati & Brennan, 2010) or inhibiting their own privileged knowledge when the task is more relevant to the addressee's perspective (Yoon et al., 2012).

However, the influence of perspective-taking on language production is more robust in production tasks involving longer, non-restrictive outputs, where participants have no constraints in word length and planning time for the utterances, such as storytelling (Galati & Brennan, 2010), offering open-ended guidance (e.g., Yoon et al., 2012), or describing unconventional, atypical stimuli that cannot be strictly modified by a limited set of property words such as in Figure 2.8 created by (Gann & Barr, 2012; for unconventional stimuli, see also Fussell & Krauss, 1989; Heller et al., 2012; Mozuraitis et al., 2018; S. Wu & Keysar, 2007).

When the utterances are more constrained by the stimuli, however, perspective-taking effects are often subtler. Some of these nuanced effects are already mentioned in the discussion above, such as in Haywood et al. (2003), Gann & Barr (2012), and Vogels et al. (2020). Divergence emerges in the extent to which perspective-taking influences speakers' production decisions and at what stage of utterance planning these decisions occur. Lockridge & Brennan (2002) argued that speakers

can incorporate the addressee's needs at the very early stage of utterance planning, revealed at the syntactic level (see also supporting evidence in children in Nadig & Sedivy, 2002). Lockridge & Brennan (2002) employed a storytelling-recall paradigm in which participants described pictures of events involving either typical or atypical instruments (e.g., stabbing with a *knife* or an *icepick*). They also manipulated the "visual copresence" of three groups of addressees. One group did not see the pictures described by the speakers, while the other two groups either shared a copy of the pictures or viewed the same pictures alongside the speakers.

The results indicated that speakers mentioned the atypical instrument significantly more frequently when the addressee had no visual access to the picture stimuli, while the mentioning rates for the atypical instrument were lower and similar across the two other addressee conditions. More notably, the atypical instrument was more likely to be mentioned early in the main clause (e.g., "Adolph stabbed the man with an icepick") rather than in a separate clause (e.g., "Adolph stabbed the man. He used an icepick"). Additionally, speakers were more likely to use an indefinite article (e.g., "an icepick" rather than "the icepick") when talking to an addressee without visual copresence.

These findings suggest that speakers may consider the addressee's perspective early in utterance planning, potentially influencing not only whether to mention an atypical instrument but also how it is syntactically integrated. Although the study did not involve a standard referential communication task or report precise details about word order in placing the atypical instrument within clauses, the descriptive results appear consistent with the idea that speakers incorporate the addressee's perspective early, both at the message encoding and grammatical encoding stages of language production. If atypical instruments can be further interpreted as high-surprisal or informative elements, this result could further hint a possible role for informativity in shaping syntactic structure.

However, another line of research argues that speakers do not routinely take the addressee's perspective during referential production. In a modified referential communication task, also known as the "director-matcher paradigm" (e.g., Keysar et al., 2000), speakers are exposed to contrastive object pairs, one of which is occluded for the addressee. For example, as shown in Figure 2.10, from the speaker's perspective, a small triangle contrasts with a big triangle, but from the addressee's perspective, only one triangle is visible. Speakers can either describe the target as "the small triangle" from their own perspective or as "the triangle" from the addressee's perspective, which is sufficient for target identification.

Studies using this paradigm have repeatedly shown that speakers often "leak" the privileged knowledge by using the more specific description ("the small triangle"), even when explicitly instructed to consider the addressee's perspective (e.g., Damen et al., 2019; Heller et al., 2012; Horton & Keysar, 1996; Keysar et al., 1998; Wardlow Lane et al., 2006; Wardlow Lane & Ferreira, 2008; Wardlow Lane & Liersch, 2012). Although speakers are aware of these differences and can clearly distinguish the referential contrasts between perspectives (e.g., Damen et al., 2019; Heller et al., 2012; Vanlangendonck et al., 2018), they often fail to adjust their utterances accordingly, possibly due to limited cognitive resources (e.g., Brown-Schmidt, 2009; Horton & Keysar, 1996). Adjustments of utterances caused by

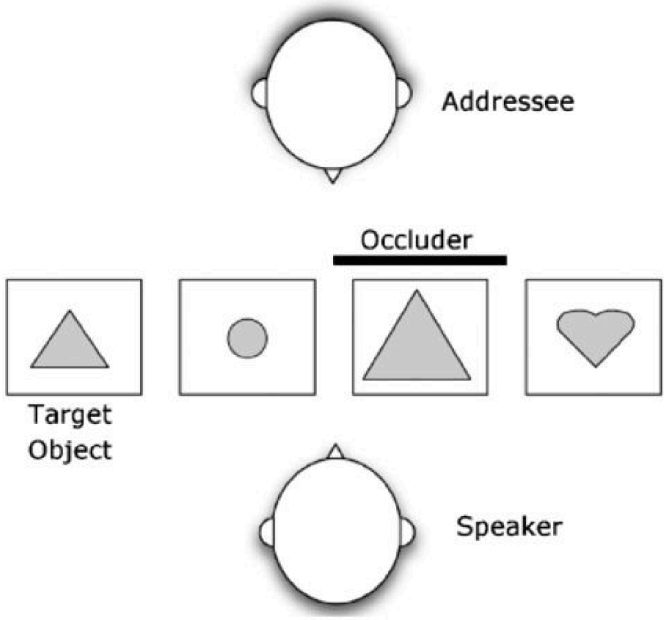


Figure 2.10: Example setup of the director–matcher paradigm from Wardlow Lane et al. (2006), Figure 1. *The big triangle* in this figure is the privileged knowledge for the speaker, as it is visible to the speaker but occluded from the addressee's view. Reproduced with permission from *Psychological Science*, © 2006 SAGE Publications. License number: 6018810657980 (issued April 30, 2025).

perspective-taking tend to occur later in the production processes or only after correction due to errors or communication failure.

It is thus still unclear to what extent perspective-taking would influence the detailed structures of speakers' utterances, such as syntactic choices. Speakers may be primarily biased by their egocentric[4] perspective (e.g., Horton & Keysar, 1996; Keysar et al., 1998), or they can incorporate the perspective of the addressee early (e.g., Lockridge & Brennan, 2002). A more moderate view combines the two extremes (e.g., Heller et al., 2012; Nadig & Sedivy, 2002; Yoon et al., 2012), but the ability to inhibit the egocentric perspective is largely dependent on the cognitive capacities of individuals, such as executive functions and working memory (e.g., Wardlow Lane, 2013).

2.2.2.5 Psycholinguistic Investigations for Post-nominal Modifications

Almost all previous studies discussed so far have focused on adjective modifiers in pre-nominal positions. Among studies investigating post-nominal modifica-

[4]I will not make further claims regarding the relationship between the egocentric perspective raised in this section and the speaker-oriented heuristic view for OS discussed in Section 2.1.1, given the qualitative differences of shared vs. privileged knowledge between the two experimental paradigms, and so the resulting categorization differences in specifications, that is, with in a set up that contains privileged knowledge visible for speakers, outputs that include adjectives are OS only from the addressee's perspective, whereas they are MS from the speaker's egocentric perspective, although one can argue that only taking the egocentric perspective is heuristic.

tions, they either leveraged cross-linguistic differences, where one language typically places modifiers after the noun (e.g., English versus Spanish speakers, as in Brown-Schmidt & Konopka, 2008; Rubio-Fernández, 2016) or use stimuli with properties that are conventionally post-nominal (e.g., an *attachment* like "a bottle with a diamond" in Figure 2.2 in Gatt et al., 2013; or an object with a *border* like "a plain in a square" in van Gompel et al., 2019). These approaches pose challenges.

The cross-linguistic comparisons are between different subject groups, meaning that the observed differences could also reflect cultural variation rather than purely linguistic effects: for example, some cultures may be more verbose in general (a concern raised by S. A. Wu & Gibson, 2021). Similarly, property-based comparisons complicate the interpretation since OS preferences vary by property type: for example, color is more frequently overspecified than size. Without knowing the baseline OS tendency for a post-nominal property, it is unclear whether speakers prefer the pre-nominal over the post-nominal structure or if they simply have a stronger preference for overspecifying certain properties, such as color, over others, such as border and attachment.

To better understand how informativity influences property ordering at the syntactic level, empirical designs should isolate ordering variation through within-subject and within-property comparisons. Even in the findings of Fukumura (2018), the results involved multinomial outcomes (i.e., the experimental outcomes involved more than two levels, see also van Gompel et al., 2019), with speakers producing both MS and OS forms as well as different ordering patterns within each condition (see Figure 2.9). This setup makes interpretation less straightforward. For instance, one could compare the increased frequency of the less preferred pattern-color order in Conditions A and B to infer an informativity effect. However, when comparing Condition A (where color is more discriminating) with Condition C, the informativity effect is less direct, as it would require summing the frequency of the color-pattern OS structure with color MS in Condition A and comparing this to color-pattern MS in Condition C. Comparing the absolute frequency of color-pattern order alone between these conditions would unfairly downplay the informativity effect, as the color-pattern order appears less frequent in Condition A due to the high occurrence of color MS. It remains unclear when and to what extent informativity affects property ordering (e.g., color-pattern OS) as opposed to property selection (color MS) in the results reported in Fukumura (2018).

Beyond these concerns about experimental design, studies on post-nominal production have identified several unique features associated with this structure. First, producing post-nominal modifications may qualitatively differ from producing pre-nominal modifications. Brown-Schmidt & Tanenhaus (2006) examined speakers' initial fixations on the contrast competitor (e.g., "the small triangle" as the target versus "the big triangle" as the competitor). These early contrast fixations reflect whether and when speakers compare the objects to determine if the adjective is needed. They observed that the timing of these fixations varied significantly between pre-nominal and post-nominal productions: while producing a post-nominal phrase, speakers' fixations on the competitor occurred not only significantly later than with a pre-nominal phrase but, in some cases, even after the onset of speaking, suggesting that speakers might initiate visual comparison until they started speaking. Brown-Schmidt & Konopka (2008) extended these findings by comparing

English and Spanish speakers' initial contrast fixations when producing descriptions with differing adjective placements based on identical referential displays (e.g., "the small butterfly" pre-nominally in English vs. "la mariposa pequeña" [the butterfly small] post-nominally in Spanish). Results showed that Spanish speakers fixated on the size competitor significantly later than English speakers.

These studies, supported by eye-tracking data, indicate that formulating a referential expression occurs incrementally by each lexical unit, rather than by initially planning the whole phrasal unit (see also Brown-Schmidt & Konopka, 2015), resulting in delayed processing of the contrastive properties when the modifiers are placed in the post-nominal position. This evidence also supports the lower rates of post-nominal color OS observed among Spanish speakers in Rubio-Fernández (2016), aligning with the Incremental Efficiency Hypothesis: when the post-nominal color in Spanish provides less necessary information, Spanish speakers prefer to conclude at the lexical unit of the noun, without further overspecifying color.

Second, post-nominal modifications may be more effortful to produce than pre-nominal ones, especially when they represent the less preferred structure in a given language. This may partially explain their overall lower frequency in production. Evidence from Fukumura & Zhang (2023) supports this interpretation: using stimuli similar to those in Fukumura (2018), they found that property ordering can be influenced by both syntactic and conceptual priming. For syntactic priming, speakers were more likely to produce a post-nominal phrase (e.g., "a striped lock that's purple") after hearing one in a preceding trial (e.g., "a dotted bowtie that's green"). They also observed conceptual priming: speakers tended to maintain the same modifier order but may switch to another syntactic structure, for example, producing "a striped purple lock" (pre-nominal) after being primed with a post-nominal pattern–color phrase. Crucially, when the less preferred syntactic or conceptual order (e.g., post-nominal or pattern–color) was not primed, speakers rarely produced it. This suggests that generating these less preferred structures may require greater production effort, making them less accessible in unprimed contexts.

Another systematic investigation into post-nominal modification looked at Hebrew, a language that favors post-nominal modifications (Trainin & Shetreet, 2021, 2023, 2024). Based on surveys of a wide range of adjective properties such as color, size, pattern, age, texture, and opinion, Trainin & Shetreet (2021) found that Hebrew speakers show a weaker preference for modifier ordering than English speakers, particularly when three adjectives are used. For instance, while English speakers consistently favor the size-color-pattern sequence pre-nominally, Hebrew speakers distribute their preferences across at least four ordering combinations for the same three properties. Notably, these combinations do not consistently adhere to the mirror-image ordering principle (e.g., Cinque, 1994): the mirrored canonical order, pattern-color-size, only accounted for about 15% of descriptions in Hebrew. These findings suggest that further research is needed to understand the factors shaping ordering hierarchies in post-nominal structures. Subjectivity and lexical frequency have emerged as influential factors (Scontras et al., 2017; Trainin & Shetreet, 2024), and Trainin & Shetreet (2023) found that listeners can recognize individual speakers by learning their unique modifier ordering patterns, indicating a communicative element.

2.2.2.6 Summary

To summarize, this section reviews psycholinguistic factors, mainly incrementality, priming, and informativity, which are central in shaping property ordering, especially in increasing the use of the less preferred orders for communicative purposes, e.g., to place the informative property earlier for efficient target search. Perspective-taking is also influential in general for referential production, but its effect on property ordering remains largely unexplored.

Notably, while these factors all contribute to property ordering, they vary in their strength of effect, especially when it comes to inverting the preferred word order. Empirical evidence so far suggests that priming is the most robust and stable factor (e.g., Cleland & Pickering, 2003; Fukumura, 2018; Goudbeek & Krahmer, 2012; Haywood et al., 2003). Other factors may require explicit instructions about word order at the outset of the experiment, such as explicitly encouraging participants to consider both orders at the outset of the experiment (Fukumura, 2018).

At the same time, depending on the specific visual context and properties involved, these factors may jointly influence speakers' output, either reinforcing or competing with each other. For instance, in the stimuli used by Fukumura (2018), accessibility would favor a color-pattern order across all conditions, while priming and informativity might predict different orders depending on the exact visual and linguistic stimuli. Future studies focusing on one of these factors should carefully identify the source of the effects and design conditions that can disentangle potentially confounding influences.

To date, few studies have focused on the production and ordering of post-nominal modifications. Current evidence suggests that producing post-nominal modifications differs qualitatively from pre-nominal structures: it is generally more difficult and less frequently used. Priming remains the most effective manipulation for eliciting a post-nominal modification when it is less preferred in the language. Additionally, modifier ordering in the post-nominal position appears less consistent, and the conventional constraints on pre-nominal adjective ordering are insufficient to fully explain the observed ordering flexibility.

In summary, while adjective ordering tends to be rigid by conventional constraints, psycholinguistic factors, particularly incrementality, priming, and informativity, play a substantial role in determining the sequence of modifiers in speakers' real-time production during referential communication tasks. These factors can increase the likelihood of the use of the less preferred ordering for communicative purposes. However, alternating between pre- and post-nominal modifier structures at the syntactic level presents additional difficulty. The current thesis project aims to examine whether this syntactic alternation can be driven primarily by the influence of informativity.

2.2.3 Information-Theoretic Accounts

A range of computational linguistic research has sought to provide quantitative, information-theoretic explanations for adjective ordering. Many studies investigate why adjectives follow specific ordering patterns, especially for their *non-restrictive use*, where adjectives describe inherent noun properties rather than distinguishing a referent within a set — contrasting with the *restrictive use* of adjectives that specify a target among referents (Hahn et al., 2018).

For *non-restrictive use*, several key predictors have emerged. For instance, *Pointwise Mutual Information* measures the likelihood of co-occurrence of adjectives and nouns (e.g., Culbertson et al., 2020; Futrell, 2019; Hahn et al., 2018); *Likely need probability* estimates reflects how strongly an adjective is associated with the semantic relevance of the noun's category (Westbury, 2021); Information gain measures how much an adjective reduces uncertainty in categorizing nouns, based on how well it separates different noun categories in a decision tree model (Dyer et al., 2021); and *syntactic category entropy* approximates integration complexity between adjectives and the head noun (Dyer, 2018). In addition to these predictors, the Uniform Information Density (UID) hypothesis (e.g., A. F. Frank & Jaeger, 2008; Jaeger, 2010) models informativity as surprisal (see Chapter 1.2.2), predicting that speakers prefer expressions that distribute surprisal more uniformly across an utterance to ease comprehension processing (e.g., Jaeger & Levy, 2006; Sikos et al., 2017). While evidence of UID is typically applied to discourse-based contexts, its implications for property ordering has been discussed in Chapter 1.4.1.

These computational metrics have managed to accurately predict adjective orders across languages by aligning the predicted results from these measures with observed corpus-based adjective orders. While these accounts may not be mutually exclusive, they all relate to concepts like surprisal and entropy, offering various forms of informativity that ultimately aim to support an efficient communicative system of language (e.g., Futrell, 2019).

For *restrictive use*, where adjectives are used to identify a target among referents, many computational models have attempted to capture the referential communication process from the perspective of the listener and/or the speaker. Among many computational accounts (e.g., models that specifically account for OS, see Dale & Reiter, 1995; Jara-Ettinger & Rubio-Fernández, 2022; Mozuraitis et al., 2018; van Gompel et al., 2019), the most relevant model to the current thesis is the Rational Speech Act framework (RSA, M. C. Frank & Goodman, 2012) that employs Bayes' rules to formalize how listeners draw pragmatic inference about speakers' intended meaning in contexts (for a systematic review, see Degen, 2023; Goodman & Frank, 2016).

The RSA model was originally evaluated in a referential communication context in M. C. Frank & Goodman (2012). A speaker should choose one word (e.g., in Figure 2.11, "blue" or "circle") to guide a listener to identify the target, and the listener should decide the target referent based on the one-word information from the speaker. In this setup, successful communication relies on both parties to make pragmatic inferences about each other rather than relying solely on literal meanings.

Figure 2.11: A recreated version of the visual context used to evaluate the RSA model in M. C. Frank & Goodman (2012), Figure A. The layout and colors were redrawn by the author to illustrate the same conceptual contrast.

The listener should be able to update their literal belief[5] about the intended referent after receiving the one-word message from the speaker, by reasoning about how likely a particular word would be adopted for a referent. For example, after receiving the word "blue" to choose between the two potential referents in Figure 2.11, a pragmatic listener would reason that if the speaker intended to refer to the blue circle, the word "circle" would be more likely to be used, as it can directly identify the blue circle, but the speaker did not say "circle", so he must mean the blue square.

A pragmatic speaker, when choosing among a set of possible words w for the intended referent r_s in a given context C, should also reason about how to convey the intended meaning clearly and efficiently to the listener. The probability of using a word w is proportional to the Utility U of the word:

$$P(w|r_s, C) \propto e^{\alpha U(w;r_s,C)}$$

where α represents the level of rationality, i.e., the extent to which a speaker can maximize Utility.

A rational and informative speaker should aim to maximize Utility U that balances between achieving maximal informativity I for the intended referent r_s, and minimizing the speaking cost D:

$$U(w; r_s, C) = I(w, r_s, C) - D(w)$$

[5]Here, "literal" shares the meaning of the "literal listener" in the RSA model, who interprets words at face value, e.g., interpreting "blue" as equally referring to any blue object in a set without considering context or the speaker's intent.

The speaking cost D represents a wide range of factors that contribute to production difficulties, e.g., word frequency, word length, accessibility, etc. M. C. Frank & Goodman (2012) assumed it constant given their simple setup (but also see a comparison of shape and color words in Sikos et al., 2021b; see also Levshina, 2022 for a more detailed discussion about cost).

If assuming constant speaking cost, maximal informativity approximates maximal Utility for speakers' rational word choices. It was quantified by surprisal $-logP(w)$, measuring the level of uncertainty reduction for the intended referents. To apply the derivation above to a specific visual context, such as Figure 2.11, the probability of using a particular word is proportional to its informativity (see Supplementary Material of M. C. Frank & Goodman, 2012 for detailed derivations).

$$P(w|r_s, C) = \frac{|w|^{-1}}{\sum_{w' \in W} |W'|^{-1}}$$

where w represents the number of objects the word w can refer to, and $|W|$ represents all possible words for the intended referent. For instance, if the target is *the blue circle* in Figure 2.11, the probability of using the word "blue" would be 1/3, while it would be 2/3 for the word "circle." This quantifies the intuition that speakers are more likely to use the more informative word "circle" than "blue" when the blue circle is the intended referent.

Although formalized differently, the probability of using a particular word in the RSA speaker model aligns with the trend predicted by the RER measure in the current project: words that refer to a smaller set of referents are more informative.

Previous research has shown that the RSA speaker model predicts human word choice fairly well under constraints that limit production to a single word (e.g., Qing & Franke, 2015; Sikos et al., 2021b). However, given the one-word production constraint and the relatively simple visual displays, it remains unclear how the RSA model would perform if extended to the production of multiple words in complex visual displays. Introducing these complications would necessitate predictions about both word choice and word ordering, aligning with what has been discussed so far in this review chapter on property selection and property ordering.

For property selection and mainly to account for varying rates of OS, Degen et al. (2020) proposed modifying the RSA model by replacing its deterministic Boolean literal interpretation with a probabilistic, continuous semantics tailored to different property types. For example, color words might be assigned a selection probability of 0.99, whereas size words might be weighted at 0.8, reflecting color's higher salience and thus greater likelihood of use. When multiple words are allowed, the probability of the property combination is calculated by multiplying the individual probabilities of selecting each property. This modified RSA model successfully predicted several empirically observed trends: higher OS rates for color than size, increased OS in visually complex scenes, and preferences for OS when objects have atypical colors. However, this model does not address property ordering, e.g., it predicts the same probability for color-size and size-color orders.

For property ordering, Cohn-Gordon et al. (2019) developed an incremental RSA model that integrates Iterated Response models, where each word choice by the speaker is shaped by both the context and prior words used for the target

referents. This process is implemented through greedy unrolling, predicting the selection of the most informative word at each stage. The model successfully accounts for the higher frequency of color OS in English in the pre-nominal position than in Spanish, as observed by Rubio-Fernández (2016). Waldon & Degen (2021) further combined the non-deterministic RSA speaker model proposed by Degen et al. (2020) with the incremental RSA model by Cohn-Gordon et al. (2019) to more effectively account for the preferred size-color order in English, as well as various word orders in Spanish for post-nominal and split structures (e.g., "a small pin that is blue"), particularly when one of the properties is overspecified.

The RSA model and its derivations emphasize the role of informativity in determining both property selection and, potentially, property ordering. They favor the most informative word when only one word is allowed, and they predict to prioritize the most informative word to appear earlier in the utterance. The hypothesis of the current research project aligns with the RSA-based models, predicting that words with a higher RER are more likely to appear earlier in the utterance (see RQ2 in Chapter 1.4.1), even when such an order may require a syntactic alternation (e.g., from pre-nominal to post-nominal modification).

To summarize this section on property ordering, previous research suggests that linearization is shaped by three main sources of influence: conventional constraints, psycholinguistic factors — primarily including incrementality and priming, but potentially also informativity and perspective-taking, as well as information-theoretic considerations. While these findings have advanced our understanding of pre-nominal adjective ordering, post-nominal modifier ordering and the alternation between pre- and post-nominal structures have received considerably less attention in the existing literature. This leaves the open question of how informativity influences syntactic linearization in referential communication, which is the central research question of the thesis. The following section outlines the detailed project design to address this question.

2.3 Project Design

By reviewing previous research on property selection and property ordering in referential expressions, this chapter has motivated the investigation of informativity in property ordering to better disentangle the speaker-oriented and listener-oriented views. This research question shifts the focus away from the predominant emphasis on property selection, particularly OS, which has shown inconclusive evidence for distinguishing between the two views.

While property ordering is typically shaped by speaker-internal pressures, such as grammatical conventions, availability and priming, this project tests whether a speaker-external, listener-oriented factor like informativity can nonetheless influence property ordering. Informativity is considered a speaker-external factor because it relates to the listener's need for efficient target identification. This identification process is assumed to unfold incrementally, with each property word helping to narrow down the referential scope (see Section 2.1.2.1 and Section 2.2.2.1). This narrowing process is quantified using RER (Tourtouri et al., 2019), which captures

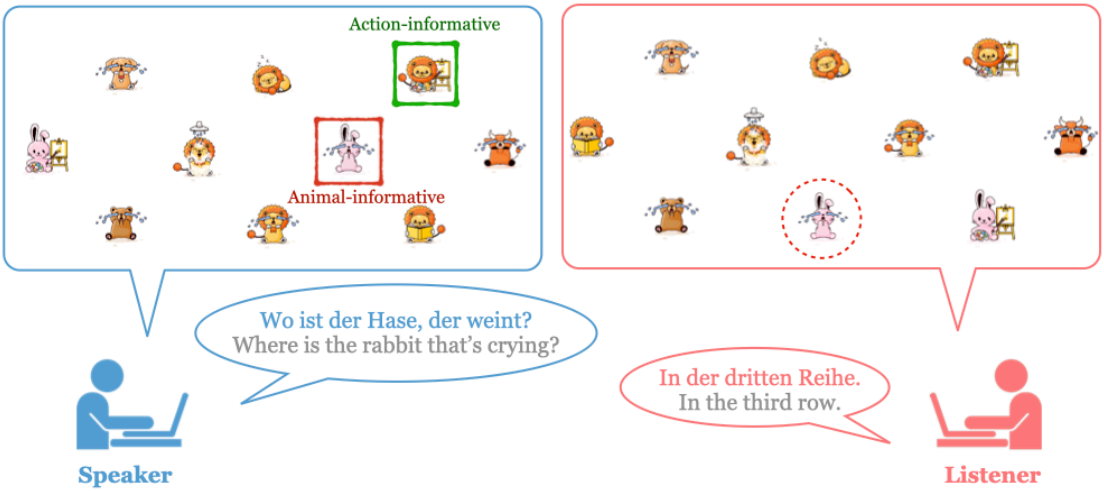


Figure 2.12: Overview of the setup for the referential communication game used in the current project. A speaker and a listener interact over a visual display containing multiple animal figures performing different actions. The animal and action properties vary in their selection scopes, differing in RER and thus in informativity. For illustration, the red and green squares mark the targets used in the main experiments for the Animal-Informative and Action-Informative Conditions, respectively. In the actual task, the speaker describes only one target per trial; in this figure, an Animal-Informative target is shown.

how much uncertainty about referring to the target referent is reduced by a given property word in a visually-situated context (see Chapter 1.2.2). Showing that informativity can modulate property ordering would provide solid support for the listener-oriented view, suggesting that speakers adapt their utterances to facilitate the listener's search and thereby enhance communicative efficiency.

At the experimental level, previous research also offered several insights that guided the experimental design of the current project.

First, property ordering tends to follow rigid, language-specific conventions that often favor pre-nominal structures. Producing post-nominal modifiers is generally more effortful and less frequent. To counter this structural bias, which might otherwise obscure potential effects of informativity on linearization, the current project adopted stimuli featuring animals performing actions (see Figure 2.12), which can be described flexibly in German (e.g., "der weinende Hase" [the crying rabbit] vs. "der Hase, der weint" [the rabbit, that cries]). The flexibility of using both structures for this stimulus set is empirically tested in the Pre-tests in Chapter 3.

Second, while informativity may help distinguish between speaker- and listener-oriented production strategies, detecting its influence on property ordering requires carefully controlled experimental manipulations. These include introducing variation in production modality (e.g., written vs. spoken), communication environments (online or face-to-face), and addressees' behavior (see Section 2.2.2.4). These factors are systematically explored in Chapter 4 and Chapter 5.

Third, as shown by Tourtouri et al. (2019), speakers may be divided in to different groups based on their sensitivity to informativity. This project investigates whether encouraging perspective-taking, specifically through speaker-listener role alternation in interactive tasks, can increase the proportion of listener-oriented speakers. Because priming effects would inevitably emerge when participants are repeatedly exposed to their partner's utterances, a confederate is introduced to both enhance perspective-taking and systematically manage the influence of priming. These aspects are implemented in Chapter 5.

Building on these insights from previous work, the current project conducted five main experiments to address the three Research Questions introduced in Chapter 1.4.1. Experiments 1–3, presented in Chapter 4, address RQ1: whether informativity influences syntactic linearization, and RQ2: how it modulates property ordering, contrasting the informative-first linearization hypothesis with UID-based predictions. Experiments 4–5, presented in Chapter 5, address RQ3: whether communication engagement modulates the effect of informativity. The design of the main experiments was first informed by two Pre-tests presented in Chapter 3, which validated the visual stimuli and linguistic candidates used in the main experiments.

Across all experiments, participants performed a referential communication task (Figure 2.12), in which they described a target referent featuring as an animal figure performing an action (encoded as a noun and a modifier, respectively). These referents could be described using either pre-nominal (e.g., "der weinende Hase" [the crying rabbit]) or post-nominal (e.g., "der Hase, der weint" [the rabbit, that cries]) structures in German, as validated in the Pre-tests.

The main experiments manipulated the informativity of either the animal or the action property through RER (Chapter 1.2.2), correlating to the selection scope of the property, i.e., the number of referents that can be modified in the visual display. Two main conditions were created: an *Animal-Informative Condition* and an *Action-Informative Condition*, in which the more informative property reduced the selection scope from ten to two figures, while the less informative one reduced it from ten to five (see Figure 2.12 for an example). Participants completed a maze-based sentence completion task in Experiments 1–4 and an interactive spoken production task in Experiment 5. Procedural details are presented in the respective chapters.

The hypotheses corresponding to RQ1-RQ3 are as follows.

For RQ1: If informativity influences syntactic linearization, the use of pre- and post-nominal modification should differ between the two conditions. Alternatively, if there is no effect of informativity, speakers would rely on the preferred syntactic structure consistently (e.g., pre-nominal modifications), regardless of conditions, as predicted by a speaker-oriented account.

For RQ2: If speakers prefer an informative-first linearization strategy, they would be more likely to place the more informative property earlier in referential expressions when comparing the two conditions, favoring pre-nominal modifications (action-animal) in the Action-Informative Condition and post-nominal (animal-action) in the Animal-Informative Condition. This prediction aligns with the listener-oriented view, which holds that speakers adapt word order to facilitate

listeners' visual search (e.g., Fukumura, 2018; Rubio-Fernández, 2016). In contrast, a UID-based linearization strategy would predict the opposite, favoring early placement of less informative content.

For RQ3: If communication engagement amplifies the effect of informativity, more engaging experimental settings and tasks should increase speakers' sensitivity to the informativity of properties and further enhance its influence on property ordering. Alternatively, if speakers exhibit similar behavior regardless of increased communication engagement, this would suggest that their sensitivity to informativity is not affected by the specific communication environments during referential production.

The following chapters present the empirical work addressing these hypotheses, beginning with the Pre-tests and continuing with the five main experiments.

3

Pre-tests: Ordering Preference in German

Contents

Chapter 2 discussed multiple factors, such as conventional constraints, priming, and informativity, that could influence the linearization of property orders. Among these, the current project focused on the effect of informativity on syntactic linearization, where alternating property orders lead to syntactic differences in referential expressions.

Before designing the main experiments to address the main research questions, it is essential to ensure that the project adopts suitable stimuli and linguistic candidates for further investigation. Specifically, the stimuli and expressions should:

1. Allow properties to be flexibly encoded in different orders within referential expressions, rather than being rigidly constrained by grammar or semantic conventions (see Chapter 2.2.1).
2. Lead to syntactic variations when properties are ordered differently.
3. Occur naturally in production, without speakers exhibiting a strong preference for one of the variants at the ceiling level, or, at the very least, allow the distribution of expressions with syntactic variations to be modulated and detected through experimental manipulations.

The Pre-tests presented in this chapter aimed to identify the visual stimuli and linguistic candidates that can meet these requirements. A promising testing ground is to compare **pre-nominal** and **post-nominal** modifications in German to describe **animals** performing different **actions** (e.g., *a sleeping dog* in Figure 3.1). Unlike simple object properties (e.g., color, pattern), which are strongly preferred as pre-nominal modifiers (e.g., Fukumura, 2018; Tourtouri et al., 2019), the action property can naturally appear in either a pre-nominal form before the animal property (e.g., "der schlafende Hund" [the sleeping dog]) or in a post-nominal relative clause after the animal property (e.g., "der Hund, der schläft" [the dog, that sleeps]) in German. This syntactic flexibility makes them suitable for testing whether and how informativity influences property ordering.

To examine whether both pre- and post-nominal structures can be naturally produced to encode the animal and action properties, and to test whether the use of the two structures is sensitive to experimental manipulations such as informativity (i.e., RER), two Pre-tests were conducted: **Pre-test 1** (Section 3.1) asked participants to describe target referents in visual contexts containing extreme RER contrasts, where the action and animal properties were either entirely uninformative or fully informative in uniquely identifying the referent. **Pre-test 2** (Section 3.2) further introduced a gradient manipulation of RER, where the informative property narrowed down a smaller selection scope than the uninformative property for target identification. The goal of Pre-test 2 is to examine whether the effect of informativity remains detectable in spoken production task when the RER contrasts between the two properties are subtler than Pre-test 1 and simultaneously compete with other factors influencing property ordering, such as consistency and visual complexity (e.g., Koolen et al., 2013; Tarenskeen et al., 2015, see also Chapter 2.1).

The Pre-tests were fundamental in addressing the broader research objectives. They ensured that the selected stimuli and linguistic candidates could be flexibly encoded in different property orders (action-animal for pre-nominal modifications, animal-action for post-nominal modifications) and that the use of the two structures can be at least modulated by extreme contrasts of RER, to be further implemented in the main experiments.

3.1 Pre-test 1

The goal of Pre-test 1, as the initial experiment of the current project, was to validate the designed stimuli depicting animals performing actions, ensuring that German native speakers could naturally produce both pre-nominal and post-nominal structures flexibly when alternating the order of the action and animal property. This would result in pre-nominal modifications in the action-animal order (e.g., "the sleeping dog") and post-nominal modifications in the animal-action order (e.g., "the dog that cries"). Additionally, the Pre-test examined whether the two property orders were sensitive to extreme contrasts in RER.

Pre-test 1 compared pre- and post-nominal modifications across two conditions with extreme RER contrasts: the Action-Informative Condition and the Animal-Informative Condition (Figure 3.1). In both conditions, the informative property

can uniquely identify the target, leading to maximal RER, while the uninformative property did not reduce entropy.

The hypotheses are, first, that both pre-nominal and post-nominal structures should be produced with detectable proportions. In particular, a reasonable number of post-nominal modifications are expected when describing this set of stimuli, if both structures are indeed allowed and flexible in German. Second, the distribution of the two structures should not be static but should vary across conditions, indicating sensitivity to experimental manipulations, particularly RER. Following the findings of Fukumura (2018) on adjective ordering, the more informative property is expected to be mentioned first more frequently when comparing the two conditions. Specifically, post-nominal modifications (starting with the animal property) might be more frequently adopted in the Animal-Informative Condition than in the Action-Informative Condition.

Participants played an online communication game and orally described target referents to help an online partner identify the target. Each condition was presented as a single trial in an experimental list. Such a single-trial design was inspired by the one-shot language games for testing the RSA framework (M. C. Frank & Goodman, 2012), with the advantage of reducing the effect of syntactic priming (M. C. Frank & Goodman, 2012; Sikos et al., 2021b).

3.1.1 Method

3.1.1.1 Participants

84 native German speakers (53 female, mean age = 24.98) were recruited online via the subject recruiting platform Prolific. 12 participants were excluded due to technical issues such as insufficient recording quality or inattentive behavior such as environmental interruptions.

3.1.1.2 Conditions

Two conditions were included: the Action-Informative Condition and the Animal-Informative Condition. In both conditions, the informative property can uniquely identify the target, reducing the selection scope from 4 figures directly to 1 figure (RER = 2 bits), while the other property was uninformative, where the selection scope remained to be 4 figures (RER=0). For example, *the sleeping dog* in Figure 3.1 is the target of the Action-Informative Condition because *sleeping* can uniquely identify the target, while *dog* does not provide useful information, given that all four figures are *dogs* in the visual display. Similarly, in the Animal-Informative Condition, *lion* can uniquely identify the target, while *reading* is not informative.

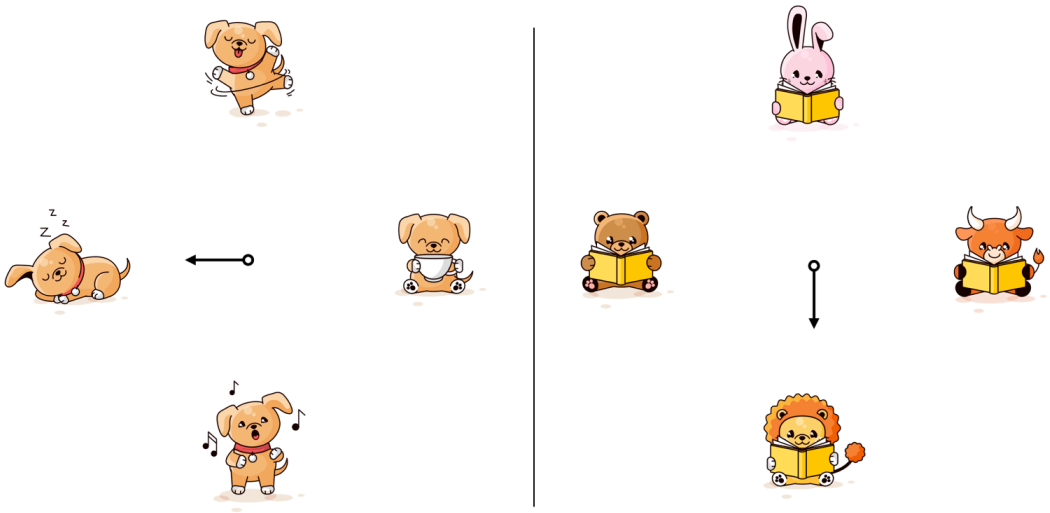


Figure 3.1: Pre-test 1 Conditions. The informative property word fully resolves referential uncertainty (RER = 2 bits), whereas the uninformative property word does not contribute to referential disambiguation (RER=0). The arrow indicates the target referent to be described by participants.

3.1.1.3 Materials, Trial Orders, and Lists

The stimuli were designed using a combination of five animals and five actions (25 individual figures in total). Images depicting these figures were created by a professional designer on commission.[1]

The five animals used as stimuli were *der Hase* [rabbit], *der Stier* [bull], *der Hund* [dog], *der Bär* [bear], and *der Löwe* [lion]. All German nouns are masculine in grammatical gender.

The five actions were *lesen* [reading], *schlafen* [sleeping], *singen* [singing], *tanzen* [dancing], and *trinken* [drinking]. The German verbs are given in their infinitive forms. These actions were selected because *sleeping*, *singing*, and *dancing* are intrinsic actions that do not involve external objects, whereas *reading* and *drinking* involve an external object in the visual stimuli (a book or a cup). This distinction was designed to test whether speakers exhibit a preference for mentioning the object in their descriptions, potentially leading to post-nominal modifications (e.g., "a dog that is reading a book").

Each experimental list contained two critical trials, one for each condition (i.e., one trial for the Animal-Informative Condition and one for the Action-Informative Condition), following a "two-shot" design. Pre-test 1 included two target sets for these critical trials: Set A (*a sleeping dog* and *a reading lion*) and Set B (*a*

[1]The designer does not retain copyright or any licensing rights; full ownership remains with the author.

drinking dog and *a singing lion*). Each target in each set was shown only once to each participant in an experimental list but was presented in both conditions across two separate lists. Additionally, the order of the two conditions in the two critical trials was reversed in two other separate lists. This resulted in four lists for each target set and a total of eight experimental lists.

The four-figure displays were generated based on the assigned target properties and the property matrix for the two critical conditions. In the Action-Informative Condition, each display featured the same animal performing four different actions, while in the Animal-Informative Condition, each display featured four different animals performing the same action. The target figure was highlighted with an arrow and displayed at the center of the screen. Three non-target figures were randomly selected from the pool of created figures based on the design of the two conditions. The positions of both the targets and the non-target figures were pseudo-randomized. The four-figure displays were generated using the R package PNG (Urbanek, 2013).

3.1.1.4 Procedure

The experiment was conducted online using the LabVanced platform (Goeke et al., 2017). It was introduced as a communication game, where participants were tasked with helping a virtual bear find its animal friend among four possible animals.

To familiarize participants with the lexicon of the properties, a practice session was conducted before the critical trials, using an audio-picture matching task. Participants clicked on five audio icons to listen to recordings that described the five figures. These recordings included both pre- and post-nominal modifications. Afterward, participants were asked to arrange the five figures in the same order as the audio recordings.

In the main experiment, each critical trial began with a four-figure display, with no indication of the target figure. After 2000 ms, an arrow appeared in the center of the display, pointing to the target figure. Participants were then instructed to directly describe the target orally (e.g., uttering "the sleeping dog" in German). Their descriptions were recorded.

To mitigate the immediate syntactic priming effect, a filler trial (a math question: 55 + 70 = ?) was inserted between the two critical trials. Participants were asked to answer this question orally. The experiment ended with a questionnaire collecting demographic information, such as participants' age, gender, and language background.

3.1.1.5 Analysis

Participants' oral descriptions were transcribed and coded into pre- or post-nominal structures. Incorrect descriptions and descriptions mentioning locations were further excluded. 7 trials were excluded in this step.

Unless otherwise specified, observed proportions reported in this thesis were computed as the number of trials showing a particular structure (e.g., post-nominal

modification) divided by the relevant total number of trials. The denominator varied depending on the comparison being reported (e.g., across all critical trials, or within a specific condition).

To accompany the reported proportions, 95% Confidence Intervals (CIs) in this project were calculated using the Clopper–Pearson method (Clopper & Pearson, 1934), a standard approach for estimating a range within which the true proportion is likely to fall. This method is based directly on the binomial distribution of binary outcomes (e.g., pre- or post-nominal), and does not rely on large-sample approximations or the assumption that the proportion follows a normal distribution.[2] It is particularly suitable when proportions are based on discrete counts, as in the current project, where each proportion reflects the number of trials using a particular structure (e.g., post-nominal modification) out of a total. The intervals were calculated using the `BinomCI()` function from the DescTools package in R (Signorell, 2025).

CIs are displayed in text format (e.g., 20% [18%, 22%]) directly on bar plots when space allows; in cases where the bar width was too narrow to accommodate the text, intervals were omitted for visual clarity. All values were computed consistently using the same method.

For Pre-test 1, a logistic regression analysis was conducted in R (version 4.3.1, R Core Team, 2023). The dependent variable was the syntactic `structures` used: pre- or post-nominal modifications. The predictor of the model was the `Condition` of the critical trials (Action-Informative or Animal-Informative Condition). Predictor `Condition` was dummy-coded, with the Action-Informative Condition as the reference level. Random intercepts and random slopes in mixed modeling (Bates et al., 2015) were not included due to the convergence issue.

3.1.2 Results

Figure 3.2 presented the overall use of the pre- and post-nominal structures in the two conditions. There was a general preference for pre-nominal modifications, starting with describing the action property (60.58% pre-nominal vs. 39.42% post-nominal modifications, collapsed across conditions). The proportions of the two modification structures used across the two conditions were similar, which were not significantly different ($\beta = -0.05, SE = 0.35, z = -0.14, p = 0.89$).

Participants did not show bias towards different types of actions, as there were only eight trials (5% of critical trials) where the external object was mentioned (a book or a cup) in the post-nominal structure.

[2]Normal distribution cannot be assumed in the current case because speakers tend to prefer the pre-nominal structure over the post-nominal one in general. As a result, the observed proportions are often skewed toward low values for the post-nominal structure. Even if sampled repeatedly, the distribution of proportions for the post-nominal structure would remain skewed rather than approximating a normal distribution.

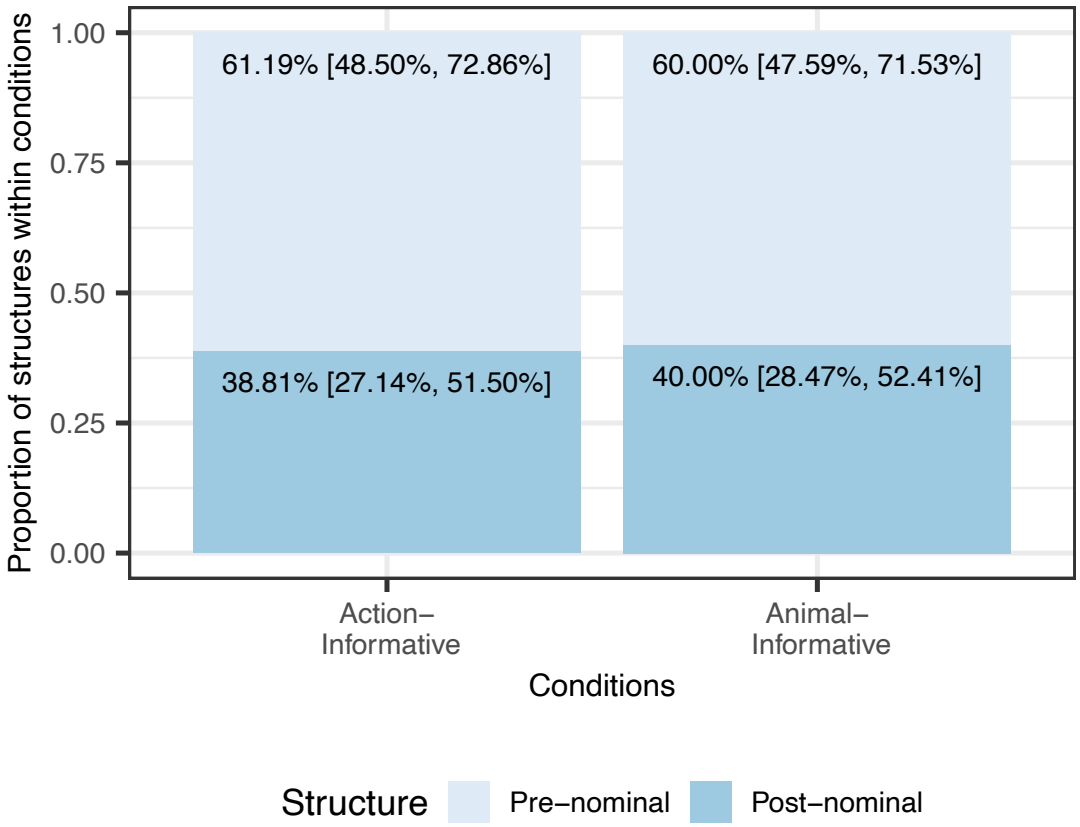


Figure 3.2: Pre-test 1. Overall proportion of the use of pre-nominal and post-nominal modifications within each critical condition. Proportions in square brackets represent 95% binomial Confidence Intervals calculated using the Clopper-Pearson method.

Speakers were consistent in the use of syntactic structures: 74 participants (88%) adopted the same syntactic structure throughout the experiment.[3] Given this consistency, the syntactic structures used in the first trial may provide more insights than the overall result. Figure 3.3 presented the proportion of the use of the two structures in the two trial orders separately. More pre-nominal modifications were adopted in the Animal-Informative Condition (71.88%, proportion within condition and trial order) than in the Action-Informative Condition (51.43%). This trend was, however, only close to significant ($\beta = 0.88, SE = 0.52, z = 1.70, p = 0.09$) based on a logistic regression analysis for data of the first trial.[4]

[3]The other ten participants alternated between pre- and post-nominal modifications across the two critical trials. They showed the same trend as the overall result, using more pre-nominal structures in the Animal-Informative condition.

[4]The close-to-significant result could be due to the small sample size (about 32 data points per condition in the first trial order) after reducing the data to half of the original sample size.

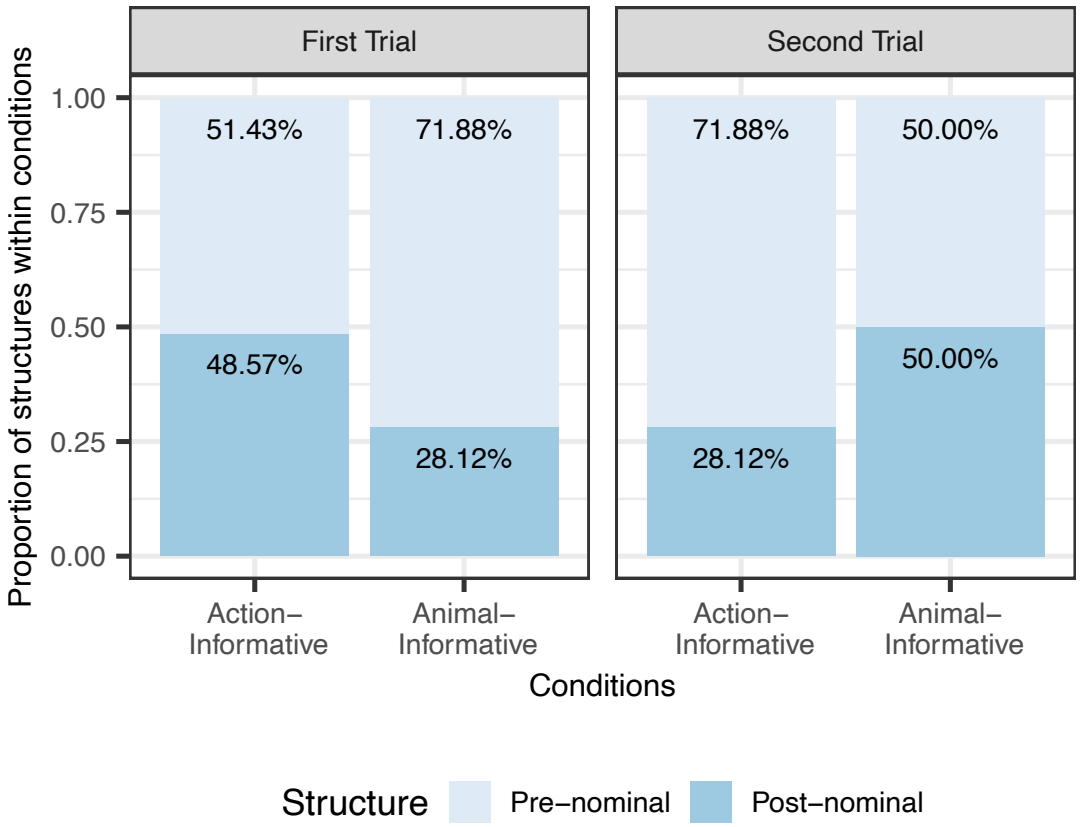


Figure 3.3: Pre-test 1. The proportion of using pre- or post-nominal structures in the two conditions across the two trial orders. Noting that the Action-Informative Condition in the First Trial and the Animal-Informative Condition in the Second Trial formed a single experimental list.

3.1.3 Discussion

To validate the visual stimuli and the flexibility of using pre- and post-nominal modifications in German, Pre-test 1 compared the use of these two structures in describing the action and animal properties across the Animal-Informative and Action-Informative Conditions. The two conditions contrasted RER in an extreme scenario: the informative property could uniquely identify the target (maximal RER), while the uninformative property provided no useful information (zero RER).

Relatively balanced use of both pre-nominal and post-nominal structures (approximately 60% vs. 40%) was observed among native German speakers for this set of stimuli, though a general preference for pre-nominal modification was noted. Additionally, focusing on the results of the first trial, where consistency or priming did not play a role, revealed that more pre-nominal modifications were used in the Animal-Informative Condition than in the Action-Informative Condition. Conversely, more post-nominal modifications were adopted in the Action-Informative Condition than in the Animal-Informative Condition.

The relatively flexible usage of the two syntactic structures, along with the proportional differences observed in the first trial, validated both the stimuli and the two syntactic structures for future experiments. This is because the use of the structures was not fixed but could be modulated by the visual contexts with detectable proportional differences. This flexibility allows for further exploration of how speakers' linearization decisions may vary when the informativity of animal and action properties change.

The result also showed that speakers tended to be consistent in their use of syntactic structures, even across just two critical trials with a filler trial in between, suggesting a consistency effect (Tarenskeen et al., 2015). Focusing on the first trial, which was not influenced by consistency, it appeared that speakers preferred to mention the less informative property first (e.g., the animal *dog* in the Action-Informative Condition, forming a post-nominal structure as shown in Figure 3.1). However, this trend might be confounded by the visual salience of the display, as the uninformative property was identical across the four figures in each condition (e.g., four *dogs* in the Action-Informative Condition). It is possible that the uninformative property was simply easier to visually identify compared to the informative and unique property within each figure. Speakers may have chosen to mention the uninformative property first because it was more visually prominent, though this remains speculative.

It was also interesting that although for the Animal-Informative Condition, a single noun (e.g., "lion") can already minimally specify the target, no such an expression was adopted by the participants. It could be due to the small sample size that minimal specifications were not captured, or due to the small number of trials of the experiment (two shots) and speakers preferred to be verbose at the very beginning of the experiment.

3.2 Pre-test 2

Pre-test 1, as described in the last section, manipulated RER to an extreme degree: one property maximally reduced entropy to uniquely identify the target, while the other provided no useful information. Pre-test 1 validated the flexibility in the use of pre- and post-nominal modifications in German when describing stimuli depicting animals performing actions, thus validating the stimuli and the two syntactic structures for future experiments. However, several concerns need to be addressed before moving on to the main experiments.

First, the observed preference for placing the less informative property earlier may be confounded by the visual salience of the displays across different conditions. In Pre-test 1, the uninformative property was identical across all figures in a condition (e.g., four *dogs* in the Action-Informative Condition), which may have been visually more salient than the unique, informative property within each figure. It is unclear how speakers would behave when the visual complexity of the displays increases, which, however, is important to address as it will be the case in the following main experiments, with a more complex RER manipulation that varies the selection scopes of different properties. A more complex scene may increase the

likelihood of speakers adopting a heuristic approach (e.g., Koolen et al., 2013), such as producing the preferred pre-nominal structure. Pre-test 2 addresses this concern by introducing a gradient RER manipulation that involves more variations in the selection scope of different properties, testing how speakers' linearization strategy would change in more complex visual contexts.

Second, it is also unknown whether the linearization strategy will change when multiple experimental trials are involved. Pre-test 1, with only two critical trials, already showed a consistent tendency for speakers to use the same structures within participants. It is unclear whether this consistency would be reinforced with additional trials, potentially making participants less sensitive to the experimental manipulation, or alternatively, speakers might exhibit greater syntactic variations once they become more familiar with the task. To investigate this, Pre-test 2 expands the number of trials to allow for repeated measures within participants, helping to assess whether the consistency effect observed in Pre-test 1 would be maintained or reduced when participants are exposed to more experimental trials.

To explore how speakers behave in extended experimental sessions with multiple trials and more complex visual scenes, Pre-test 2 manipulated RER in a gradient manner, employing repeated measures within participants. One property consistently reduced entropy by half, narrowing the selection scope from four figures to two, while the other property was either informative (narrowing the scope from four figures to one, leading to unique target identification) or uninformative (narrowing the scope from four figures to three, resulting in greater referential uncertainty). This design aimed to examine how gradient RER influences speakers' linearization decisions. Similar to Pre-test 1, participants engaged in an online communication game with a simulated partner, where they helped locate target referents within these varied visual contexts.

3.2.1 Method

3.2.1.1 Participants

81 native German speakers were recruited online via Prolific. 11 participants were excluded due to technical issues or inattentive behavior. 70 participants (37 female, mean age = 33.5) were included in the final analysis.

3.2.1.2 Conditions

Critical Conditions Pre-test 2 included four critical conditions (Figure 3.4), which systematically varied the RER of the animal and action properties. Each condition was defined by the interaction between a **constant property**, which always had equal RER, and the informativity of the **variable property**.

The **constant property** was defined as the property that was always shared by two of the four figures in each display. This ensured that the corresponding property word consistently reduced entropy by 1 bit ($log_2 4 - log_2 2$) and remained unchanged across the Informative and Uninformative conditions (illustrated in each row of Figure 3.4). The **variable property**, the one with varied RER in the **Informative** versus **Uninformative** Conditions, was either unique among the

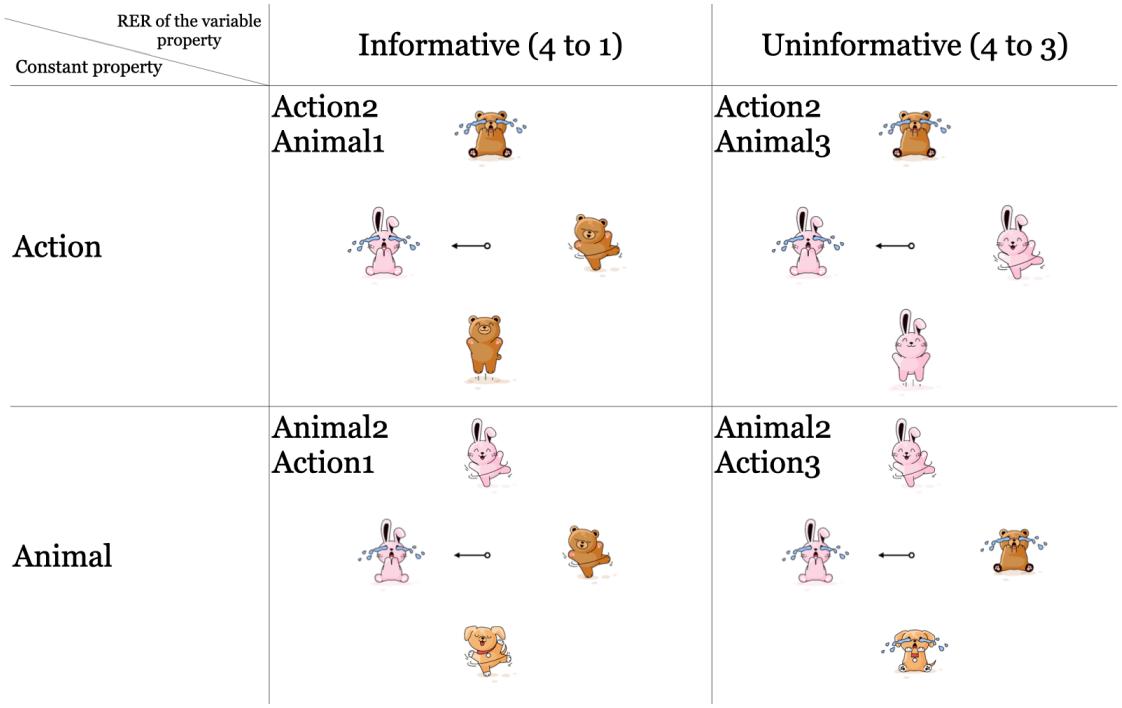


Figure 3.4: Pre-test 2. Critical Conditions. While one of the properties reduces entropy to half (reducing from 4 to 2 figures), the other property can be either informative (4 to 1 figures) or uninformative (4 to 3 figures). The name of each condition is displayed in the top-left corner of each cell, indicating the number of figures in the display sharing the target's action and animal properties (e.g., Action2Animal1 reflects two figures sharing the targets action and one sharing its animal property).

four figures, making the corresponding word informative by reducing entropy by 2 bits ($log_2 4 - log_2 1$), or shared among three figures, making the word uninformative by reducing entropy by 0.4 bits ($log_2 4 - log_2 3$).

For example, in the first row of Figure 3.4, the target figure is *the crying rabbit*. The constant property is the action *crying*, as there are always two *crying* figures in both the Informative and Uninformative Conditions. Within this set, the variable property, the animal property, varied in its informativity in RER: in the Informative Condition, only one *rabbit* is present, making the word "rabbit" unique and informative for target identification; in the Uninformative Condition, three *rabbits* are included, making the word less useful and informative for distinguishing the target.

Filler Conditions Pre-test 2 included filler trials between each critical trial to reduce the tendency for consistent syntactic choices. Additionally, it helped to mitigate the visual repetition of four-figure displays by presenting 2 or 3 figures randomly positioned at the four corners of a diamond-shaped display (see Figure 3.5 for an example).

The filler trials were designed to elicit either pre-nominal or post-nominal structures by requiring participants to distinguish targets based on *size* (e.g., "a big fox" or "a small ram" as pre-nominal modifiers) or *surface* properties (e.g., "the fox on a computer" or "the fox on a sweater" as post-nominal modifiers). These

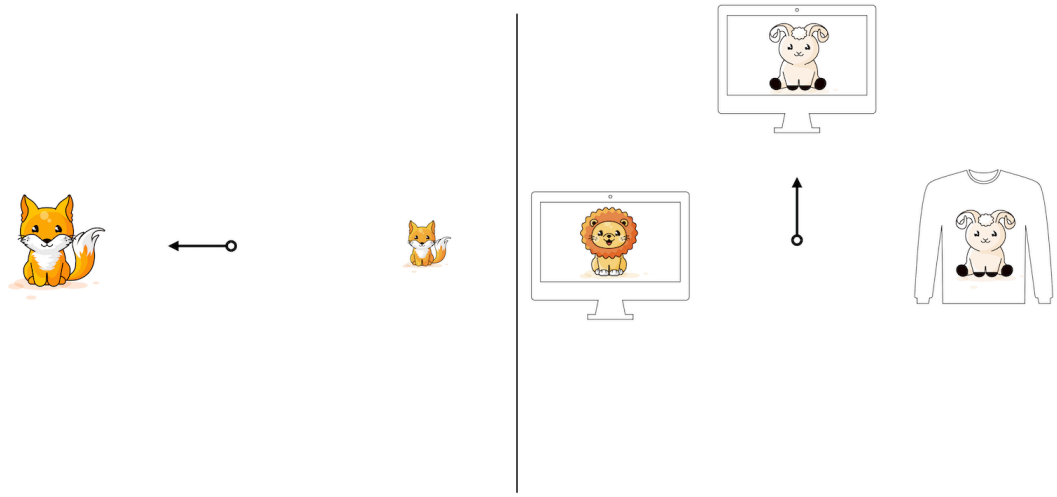


Figure 3.5: Pre-test 2 Filler Conditions. A pre-nominal or post-nominal modifier is mandatory to uniquely identify the target.

modifiers were mandatory for target identification. To avoid lexical repetition, the fillers featured a distinct set of animals (*ram*, *fox*, *rooster*, and *lion*) that did not overlap with those used in the critical conditions.

3.2.1.3 Materials, Trial Orders, and Lists

The stimuli for the critical trials were constructed using four animals: *Hase* [rabbit], *Stier* [bull], *Hund* [dog], and *Bär* [bear], and eight actions: *duschen* [shower], *springen* [jump], *weinen* [cry], *singen* [sing], *trinken* [drink], *malen* [paint], *schreiben* [write], and *tanzen* [dance].

While Pre-test 1 included only four actions, Pre-test 2 expanded the set to eight to examine whether a broader range of action words in German would still allow flexibility in pre-nominal and post-nominal modifications. The images were created by the same designer as in Pre-test 1.

A total of eight experimental lists were created by combining four lists of the critical conditions with two lists of the filler conditions. Each experimental list contained 16 trials: 8 trials for the critical conditions and 8 trials for the filler conditions. The critical trials included 8 unique target figures, created by randomly pairing each of the four animals with two different actions, ensuring that each action appeared once in a list. Four critical trial lists were generated using a Latin-square design to ensure that each target figure appeared in all four critical conditions across the experiment but only once per participant. The four-figure displays were constructed following the same procedure as in Pre-test 1, structured according to the properties of the target figures and the property matrix for the four

critical conditions. The target figure was highlighted with an arrow and displayed at the center of the screen. The location of the target figures was randomized and counterbalanced.

The two filler trial lists each contained four pre-nominal fillers and four post-nominal fillers. To prevent reinforcing the same structure, they were arranged in alternating pairs, with one pre-nominal and one post-nominal filler forming a group. The two filler lists differed only in the order of these pairs, ensuring that each critical trial was equally preceded by both types of filler trials across lists.

Each experimental list was then formed by combining one of the four critical trial lists with one of the two filler trial lists, resulting in a total of eight experimental lists. To further reduce the immediate priming effect of the filler trials on the subsequent critical trials, a math question was inserted between the filler and critical trials. Participants were instructed to answer the math question orally.

3.2.1.4 Procedure

The experiment was introduced as a communication game, where the participants were guided to collaborate with a partner connected via the internet. The participants should find out whether the location of the target figure was the same as their partner's.

Before the main experiment, the practice session familiarized participants with both communication roles: speaker and listener. As listeners, they responded with "yes" or "no" after hearing a question about the target's location. As speakers, they asked about the target's location following the same procedure as in the main experiment. After the practice session, participants were informed that they had been randomly assigned the role of the speaker in the main experiment, while their partner, who was simulated by the experimenter, would take on the role of the listener.

In the main experiment, each trial began with a four-figure display, with no indication of the target. After 2000 milliseconds, an arrow appeared in the center of the display, pointing to the target figure. Participants could then begin speaking by asking, "befindet sich _____ (target) _____ (location)?" [Do you find ____ (target) on the ____ (location)?] (e.g., "Befindet sich der weinende Hase links?" [Do you find the crying rabbit on the left?]). Participants' oral descriptions were recorded.

Once participants finished speaking, they were instructed to press the Spacebar on the keyboard to stop the recording. After a 1000-millisecond delay, the arrow in the center of the screen was replaced by a speech bubble containing two ellipsis marks ("....."), signaling that the listener, simulated by the experimenter, was processing the question. After approximately 1000 milliseconds (with some jitter, between 800 and 1000 milliseconds), the ellipsis marks were replaced by the predefined response from the partner, either "yes" or "no." In one-quarter of the trials, a "no" answer was given, indicating that the target location was different from the listener's perspective.

Table 3.1: Pretest 2. Statistical results from the logistic mixed-effects model examining the effect of Condition on Structures.

Predictor	$\hat{\beta}$	95% CI	z	p
Intercept	-3.85	[-4.84, -2.85]	-7.59	$< .001$
Constant property: Action2 vs. Animal2	0.02	[-0.81, 0.85]	0.05	.960
Action2Animal1 vs. Action2Animal3	-0.02	[-1.23, 1.18]	-0.04	.969
Animal2Action1 vs. Animal2Action3	-0.56	[-1.71, 0.60]	-0.94	.345

Note. The model was specified as: Structure Condition + (1|subject). Condition was sum-to-zero coded. The first contrast compares (second row of the table) the two constant-property conditions (Action2 vs. Animal2). The second and third contrasts (third and fourth row of the table) compare the Informative and Uninformative conditions (1 vs. 3) within each constant-property pair.

3.2.1.5 Analysis

Participants' oral descriptions were transcribed into texts and coded into pre- or post-nominal modifications. Incorrect descriptions and descriptions with repair were further excluded. 34 trials were excluded in this step.

Proportional data and their associated CIs were calculated in the same way as in Pre-test 1 (see Section 3.1.1.5).

A logistic mixed-effects regression analysis was conducted using the `lme4` R package (Bates et al., 2015). The dependent variable was the syntactic `Structure` produced by participants, with two levels: pre-nominal and post-nominal modification. The main predictor was `Condition`, which had four levels corresponding to the critical trial types. `Condition` was sum-to-zero coded to enable specific comparisons: the model first contrasted the two constant property types (i.e., the two rows in Figure 3.4) and then compared the Informative and Uninformative Conditions within each property type (within each row). The model included a random intercept for subjects. Due to convergence issues, additional random slopes or item-level random effects were not included.

3.2.2 Results

Pre-nominal modifications were overwhelmingly preferred across all four conditions (93.35% pre-nominal vs. 5.32% post-nominal modifications, collapsed across conditions). A single-noun structure (e.g., "the rabbit") was produced only seven times, all of which occurred in `Action2.Animal1` condition (see Figure 3.4), forming minimal specifications that can uniquely identify the target figure.[5]

Figure 3.6 illustrates the overall distribution of pre-nominal and post-nominal structures across conditions. The distribution of structures across conditions did not differ significantly, as summarized in Table 3.1.

[5]These seven data points were excluded for the logistic mixed model regression analysis as the outcome should be binary for the model.

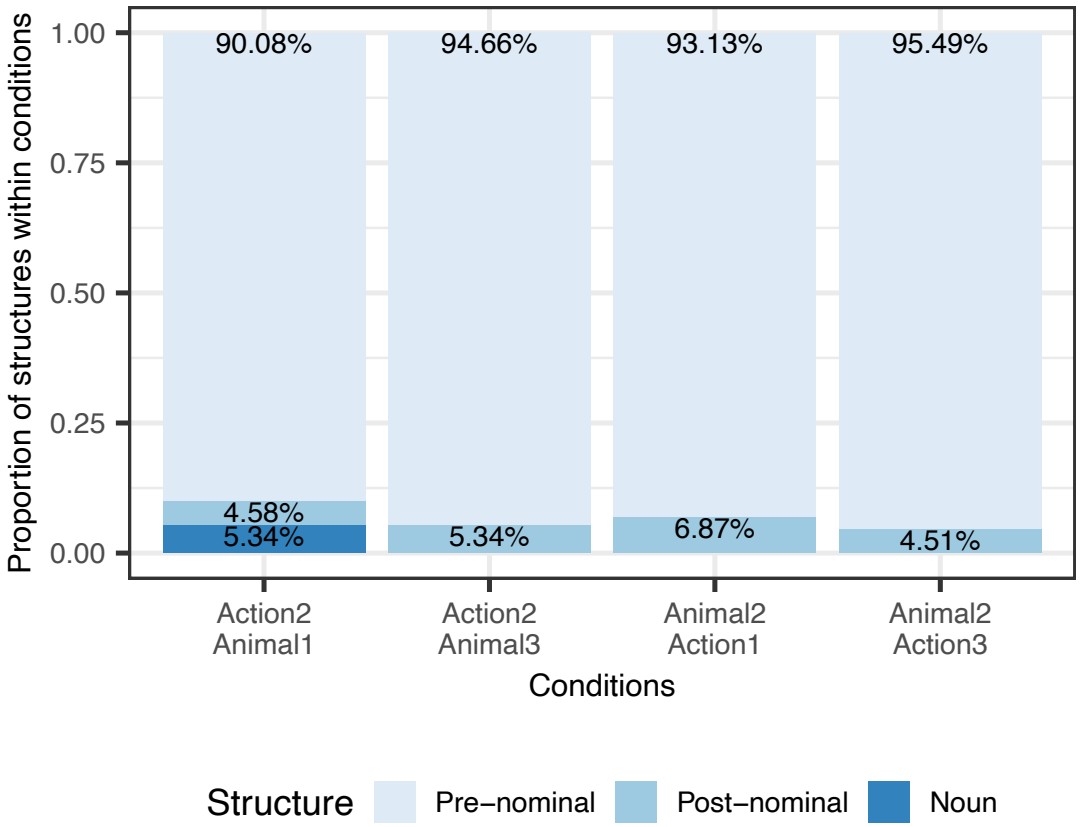


Figure 3.6: Pre-test 2. Overall proportion of the use of pre-nominal and post-nominal modifications within the four critical conditions.

3.2.3 Discussion

To further validate the stimuli and explore the flexibility of pre- and post-nominal structures when describing visually complex displays across multiple experimental trials, Pre-test 2 varied RER in a gradient manner across four conditions. However, the results revealed an overwhelming preference for pre-nominal modifications, regardless of RER. Also, minimal specifications were rarely adopted, though they were possible in `Action2.Animal1` Condition. These results suggest that speakers recruited via the internet did not significantly vary their syntactic choices based on the gradient change in RER during oral production when the visual scenes were more complex. This result is further discussed and compared with the result of Pre-test 1 below for general discussion.

3.3 General Discussion

This chapter introduced two Pre-tests to examine speakers' preferences and flexibility for pre-nominal and post-nominal modifications in German referential expressions, adopting a set of visual stimuli depicting animals performing actions. Both

experiments also investigated whether these syntactic choices could be modulated by the visual contexts, especially the differences in RER, which was operationalized through variations in the number of properties available to narrow down the selection scope in the visual displays.

While the two Pre-tests shared this overarching goal, they differed in how RER was manipulated. Pre-test 1, using a "two-shot" design, tested conditions with extreme RER contrasts, in which one of the properties was either maximally informative to uniquely identify the target, while the other was uninformative. Pre-test 2, using repeated measures with extended experimental lists, implemented a more gradient manipulation of RER, comparing properties that could be either highly informative (larger RER) or less informative (smaller RER).

Pre-test 1 demonstrated that both pre-nominal and post-nominal structures were flexibly used by native German speakers despite an overall preference for pre-nominal modification. Importantly, the presence of a reasonable amount of post-nominal structures confirmed that the selected syntactic forms and stimuli were appropriate for studying syntactic linearization in referential expressions. The result also revealed that the use of the two structures was not fixed but appeared to be sensitive to visual contexts: in the first trial, which was unaffected by syntactic priming, the uninformative property was more likely to be mentioned first when comparing conditions, suggesting that informativity might influence linearization. However, this trend may also be confounded by visual salience: the uninformative property was repeated across all figures, possibly making it more visually prominent and thus more likely to be mentioned first. Whether speakers prioritized the less informative property because of its salience or because of its syntactic function remains open to interpretation.

Pre-test 2 explored whether more subtle differences in RER could modulate syntactic structures across multiple trials. Unlike Pre-test 1, the results revealed a strong dominance of pre-nominal modifications across all conditions. One possible explanation is that the increased visual complexity of the displays in Pre-test 2, introduced to meet the requirements of the gradient RER manipulation, may have reduced speakers' sensitivity to informativity. Specifically, the inclusion of more diverse visual elements (e.g., more complex contrasts in colors and properties than in Pre-test 1) could have made the scenes cognitively more demanding, prompting speakers to rely on the preferred pre-nominal structures as a heuristic approach.

This heuristic strategy is in line with findings from prior studies (e.g., Gatt et al., 2013; Koolen et al., 2013), which reported higher rates of overspecification when speakers described targets in displays that are visually more complex and consistently favor pre-nominal modifiers regardless of the discrimination power of properties. While the present study did not investigate overspecification directly, the strong preference for pre-nominal structures in Pre-test 2 may reflect a similar influence of visual complexity. It is possible that, under increased visual load, speakers relied on the preferred syntactic forms, but ignoring the visual differences of selection scopes and RER when linearizing referential properties.

Reflecting on the findings from both Pre-tests, certain improvements are necessary for future experimental designs. One key consideration is the need to counterbalance visual factors, especially visual complexity and visual salience, across

conditions when manipulating RER. Both visual factors may be plausible to account for some of the results observed in Pre-test 1. For example, the higher frequency of pre-nominal modifications in the Animal-Informative Condition might be explained by the greater visual complexity of the displays, which featured four different animals with varying colors, leading to the tendency to use the preferred, pre-nominal structure. In contrast, the Action-Informative Condition displayed the same animal figure across all four figures, potentially making the scene visually less complex. Alternatively, the result of Pre-test 1 could also be accounted for by visual salience, where the uninformative property was repeated across the four figures and thus may have been more likely to draw speakers' attention and be mentioned first.

It is also worth considering that visual salience is not independent of RER but is inherently tied to it. A less informative property (i.e., one with a lower RER) corresponds to a broader selection scope, meaning that more figures in the display share that property. As a result, this property may become more visually prominent but less helpful for uniquely identifying the referent. Addressing the potential confound of visual salience will require careful counterbalancing across conditions, such as assigning a property to be uninformative to both the animal and action property across different conditions.

In contrast, visual complexity can be more directly controlled and kept consistent across conditions. The experiments presented in the following chapters address this by adopting a ten-figure display with a fixed level of visual complexity across all conditions. This design aims to more cleanly isolate the effect of RER while minimizing confounding influences from visual complexity.

Another consideration concerns the size of the RER contrast between the two properties. In the current design, the difference in the selection scope across properties may have been too subtle to be noticed by speakers. For example, in Pre-test 2, the more informative property differed from the constant property by only one figure (e.g., narrowing the selection scope from two to one figure), which corresponds to a relatively small RER difference (less than 1 bit). It is possible that this contrast did not provide a strong enough cue for speakers to invest additional effort in strategically selecting one syntactic structure over the other. A similar concern has been raised about the stimuli used in Gatt et al. (2013) (see Chapter 2.1.1), where limited contrast in discrimination power may have reduced the likelihood of detecting an effect on property selection.

To address this issue, the experiments presented in the following chapters increase the contrast in RER between the animal and action properties. This adjustment is intended to make the more informative property more useful for identifying the target by narrowing down a larger selection scope, thereby increasing the likelihood of detecting its influence on property ordering with syntactic choices. By creating a greater RER distinction, the design aims to assess more effectively whether speakers adapt their linearization strategies in response to differences in informativity.

Finally, relying on speakers' oral production outputs in online experiments as the primary measure of informativity effects may offer a relatively coarse-grained assessment. Various factors influencing the production process may obscure the effect of informativity itself. For instance, verbal production is highly sensitive to

the priming effect, which can lead to the consistent use of a particular syntactic structure regardless of conditions, as observed in Pre-test 2. Moreover, online participants may not consistently attend to the broader visual contexts, adopting a more heuristic, speaker-oriented strategy by focusing solely on the target figure rather than comparing it to other figures that vary in RER. As a result, the manipulation of RER may be overlooked by participants.

In addition, the presence of a simulated, hypothetical listener may not elicit the same communication dynamic as interacting with a physically present addressee. Previous research has shown that speakers are often less listener-oriented when the addressee is imagined or simulated (e.g., Ferreira et al., 2005; Kuhlen & Brennan, 2013), which may limit the influence of speaker-external factors on property selection and ordering.

Given these limitations, future experiments require more fine-grained measures of speakers' linearization decisions and a more involving experimental tasks. To address this, Experiments 1-4 in the following chapter employed a maze-based sentence completion task, which allowed for a more precise assessment of the first property selected for expression. The online partner, who was simulated in Experiments 1-4 as online experiments, was also made more authentic and involved in the communication task, by having it providing feedback and auditory descriptions.

In conclusion, the Pre-tests conducted in this chapter provided validation for both the selected visual stimuli and the use of pre- and post-nominal modifications as suitable materials for investigating the effect of informativity on property ordering for future experiments. The results confirmed that native German speakers can flexibly produce both syntactic forms when describing visual figures depicting an animal performing an action. Although a general preference for pre-nominal structures was observed and the characteristics of the visual displays may also affect the preferred syntactic structure, the proportional distributions of the two syntactic structures can be modulated by the visual contexts, as in Pre-test 1, indicating that the use of the two structures is responsive to experimental manipulations. These findings show positive evidence for the implementation of this stimulus set and syntactic structures in the main experiments that follow.

4

The Role of Informativity in Syntactic Linearization

Contents

Chapter 3 introduced two Pre-tests that validated a set of visual stimuli depicting animals performing actions to be adopted in the main experiments, as well as the use of pre- and post-nominal modifications in German. The results of the Pre-tests confirmed that both the stimuli and the syntactic structures were suitable for the main experiments investigating the effect of informativity on syntactic linearization. However, the Pre-tests also highlighted the need to better control for

visual complexity and salience across conditions and to adopt a more fine-grained measure of speakers' linearization decisions.

Building on these findings, this chapter presents three experiments designed to address RQ1 and RQ2 of the current project (see Section 1.4.1): whether informativity modulates property ordering at the syntactic level, and if so, which order is preferred.

These questions were investigated by comparing two critical conditions in three experiments introduced in this chapter (Figure 4.1): **the Action-Informative Condition** and **the Animal-Informative Condition**. Both conditions involved visual displays with ten animal figures performing different actions. The animal and action properties differed in their selection scopes, which in turn differed in the RER of the corresponding property words. The informative property word led to a larger RER, narrowing the selection scope from 10 to 2 figures, compared to the uninformative property, which narrowed the selection scope from 10 to 5 figures. The visual displays used in the main experiments expanded the contrast in RER from the Pre-tests, ensuring that the salience and complexity of the two properties were counterbalanced across conditions, with both critical conditions using identical display structures.

All three experiments also employed a **maze-based sentence completion task** ("**Maze Task**" below) to record speakers' referential expressions and so their linearization decisions. The procedure for this task will be described in detail in Section 4.1.1.4. By varying the number of word options based on the selection scope of each property in the display, the task emphasized the RER dimension, encouraging participants to consider the informativity of the properties in visual displays. Additionally, the Maze Task presented both pre- and post-nominal syntactic structures as explicit options for participants to consider both structurs for descriptions. This task allowed for a more direct and precise measure of speakers' linearization choices by capturing the first property selected in target descriptions.

Two alternative hypotheses are considered across the three experiments: First, if the use of the two syntactic structures can be modulated by RER across the two conditions, it would support RQ1 of the thesis, suggesting that informativity does influence linearization at the syntactic level. Additionally, the proportional distribution of the two structures across conditions would provide insights for RQ2, specifically regarding which property order is preferred in response to informativity. Alternatively, if speakers' ordering preferences remain consistent in favor of the pre-nominal structure, regardless of the conditions and the Maze Task design, this would suggest that informativity does not influence linearization at the syntactic level.

Experiment 1, described in Section 4.1, initiated this investigation by having participants perform the role of speakers using the Maze Task in an online communication game with a simulated listener. Experiments 2 and 3, presented in Sections 4.2 and 4.3, enhanced participants' involvement in the tasks, aiming to amplify the effect of informativity beyond what was observed in Experiment 1. To increase the sense of communication with the online partner, Experiment 2, before the main Speaker Maze Task, first asked participants to perform a Listener Task block, which engaged them in target identification tasks guided by auditory target

descriptions that prioritized the informative property in ordering. Experiment 3, building on Experiment 2, further animated the visual displays by showing the figures in three consecutive groups. This design aimed to encourage participants to make a more thorough visual preview of the displays before describing the target.

4.1 Experiment 1

As the initial experiment adopting the Maze Task and comparing RER contrasts in ten-figure displays, Experiment 1 explored whether syntactic linearization can be modulated by informativity at all. In Experiment 1, participants communicated with an online, simulated partner to determine whether they both viewed the same or a different visual display. Participants performed the role of the speaker and used the Maze Task to inquire about the location of the target. Their online partner, acting as the listener, provided feedback regarding the location of the target from the partner's perspective. Following this, participants assessed whether the displays were identical or different by comparing the target's location from their own perspective and the listener's response.

4.1.1 Method

4.1.1.1 Participants

80 participants (33 female and 47 male) were recruited online via the subject recruiting platform Prolific. Participants were native German speakers with an age range from 18 to 55 (mean = 28.73). Participants were paid according to the German minimum wage (13.94 euro/hour).

4.1.1.2 Materials

The stimuli were selected from the pool of individual figures created in the Pre-tests, combining five animals and five actions. The five animals were *der Hase* [rabbit], *der Stier* [bull], *der Hund* [dog], *der Bär* [bear], and *der Löwe* [lion]. The five actions were *duschen* [shower], *weinen* [cry], *lesen* [read], *malen* [paint], and *schlafen* [sleep]. These figures did not show a strong preference for any particular syntactic structure in the Pre-tests and were easily identifiable by the participants, with minimal variations in how the action or animal properties were named.

The 25 figures were then combined into ten-figure displays (Figure 4.1 as an example), generated using the R package `PNG` (Urbanek, 2013). Each display consisted of ten figures, with overlaps between two animal properties (e.g., 2 *rabbits* and 5 *lions* in Figure 4.1) and between two action properties, (e.g., 2 *painting* and 5 *crying* animals). This display structure ensured that each display could be used in both the Animal-Informative and Action-Informative Conditions (see Section 4.1.1.3 below), allowing control over visual complexity. The ten figures in the display were arranged in three separate rows: the top and bottom rows each contained three figures, while the middle row contained four figures.

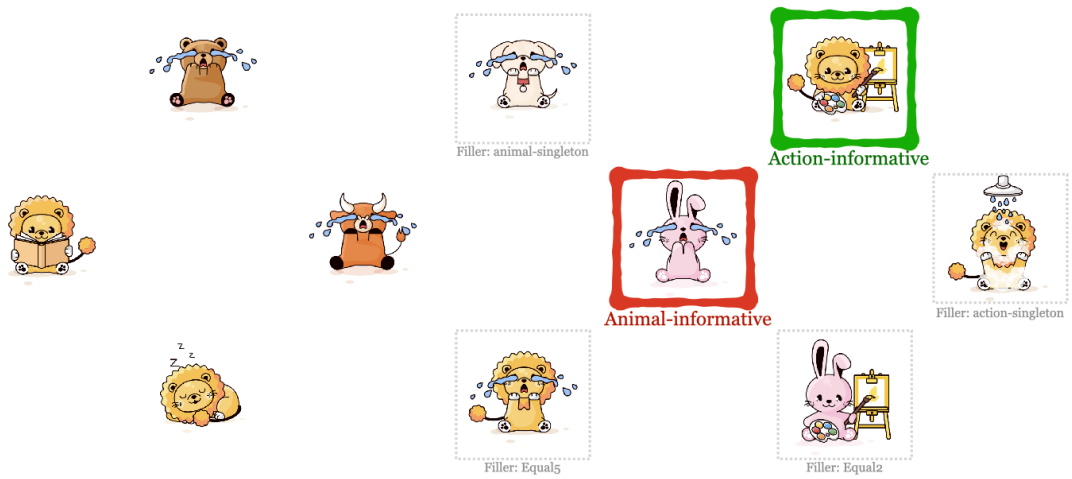


Figure 4.1: Example stimuli for conditions in Experiments 1-5: In the Action-Informative Condition, the example target figure is *the painting lion* in the green square, while in the Animal-Informative Condition, the target is *the crying rabbit* in the red square. Filler targets are marked with grey squares (Equal 2 and Equal 5 were used in Experiments 2 and 3). Only one target is highlighted in a red square during the experimental trials.

4.1.1.3 Conditions, Lists, and Trial Orders

Critical conditions The main experiments in the current project focus on comparing two critical conditions: the Animal-Informative Condition and the Action-Informative Condition. In each condition, the informative property of the target reduces more entropy compared to the uninformative property. For example, in the Animal-Informative Condition, the target is *the crying rabbit* (Figure 4.1). Given that there are only two *rabbits* in the display but five *crying* animals, the property word "rabbit" is more informative than "crying". "Rabbit" reduces the selection scope from 10 to 2 figures (RER = 2.3 bits), while "crying" only reduces it from 10 to 5 figures (RER = 1 bit). In this context, "rabbit" provides more certainty and is thus more informative than "crying." In the Action-Informative Condition, the target is *the painting lion* (Figure 4.1), where the word "painting" (reducing the scope from 10 to 2 figures) refers to a smaller selection scope compared to "lion" (from 10 to 5 figures). Therefore, the action word "painting" is more informative in this condition.

Filler conditions To reduce the tendency of consistency and priming (Pickering & Garrod, 2004; Tarenskeen et al., 2015), Experiment 1 adopted two filler conditions, the Animal-Singleton Filler and the Action-Singleton Filler, to increase the variability of linguistic structures and visual targets. A filler target contained a unique animal or action property among the ten figures in a display, which means that only using the corresponding word of the unique property can already minimally specify the target. For example, *the crying dog* in Figure 4.1 is a target of the Animal-Singleton Filler Condition, as the word "dog" can uniquely identify the target among the ten figures, while *a showering lion* is a target of

the Action-Singleton Filler Condition where "showing" can uniquely identify the showering lion.

Lists and Trial Orders In total, four experimental lists were created by combining two lists of critical trials and two lists of filler trials. Each list contained 48 trials: 24 critical trials (12 trials for the Animal-Informative Condition and 12 trials for the Action-Informative Condition) and 24 filler trials (12 trials for the Animal-Singleton Filler and 12 trials for the Action-Singleton Filler).

For the critical trials, 24 displays were created, with each display containing the two targets for both conditions. The targets and so the two conditions of each display were presented separately across the two lists. For example, *the crying rabbit* in the Animal-Informative Condition in Figure 4.1 was presented in one list and *the painting lion* in the Action-Informative Condition in the same display was presented in the other list.

At the same time, the conditions in which each target figure was assigned were also counterbalanced across the two critical lists. For example, *the crying rabbit* was the target for the Animal-Informative Condition in one list and it was the target for the Action-Informative Condition in a different list in a different visual display. Within each list, the 24 target figures were unique, ensuring no visual repetition. The distribution of animal and action properties was also counterbalanced across conditions. Within each list (24 trials), every animal and action property (from a set of five) was presented at least twice.

The position of the target within a display was pseudo-randomized across the ten figure positions in the three rows, with each position featuring a target at least twice within the 24 displays. The order of the 24 critical trials was also pseudo-randomized to ensure that both conditions were evenly distributed between the first and second halves of an experiment.

For the filler trials, 24 additional displays were created and counterbalanced similarly to the critical trials. Each filler trial preceded a critical trial, and to avoid lexical repetition, no target properties were repeated between adjacent filler and critical trials. Furthermore, the conditions of the filler trials alternated between the critical trials: the Action-Singleton Filler always preceded the Animal-Informative Critical Condition, while an Animal-Singleton Filler always preceded the Action-Informative Critical Condition. This alternation aimed to 1) obscure the patterns in the critical conditions, which consistently presented a visual contrast between two and five figures, and 2) potentially reduce the effect of consistency (Tarenskeen et al., 2015) by introducing qualitatively different linguistic structures in the filler trials for the Maze Task (introduced below).

4.1.1.4 Maze-based Sentence Completion Task

Experiment 1 employed a maze-based sentence completion task inspired by sentence comprehension studies (e.g., Forster et al., 2009). The task captured participants' choice of the first visual property as a function of informativity in the two critical conditions.

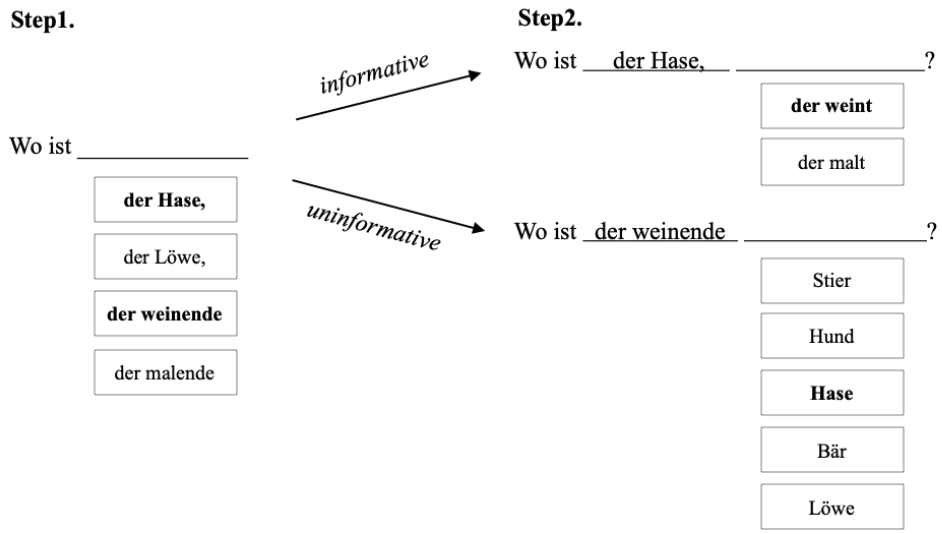


Figure 4.2: Example of the maze-based completion task for the critical conditions: The target was *the crying rabbit* in Figure 4.1. In German, it can be described in either a pre-nominal structure, "der weinende Hase" [the crying rabbit] or in a post-nominal relative clause, "der Hase, der weint" [the rabbit, that cries]. The two steps were presented sequentially. Only one stem was presented in Step 2 for participants, depending on their decisions in Step 1.

The Maze Task consisted of two steps, where participants were required to fill in two blanks consecutively (see Figure 4.2). In **Step 1**, four linguistic candidates were presented: two options matching the properties of the target figure and two distractors. These options were displayed as buttons for mouse-clicking stacked vertically in a pseudo-randomized order. The only constraint was that options with the same part of speech (noun or adjective) were adjacent for a clear visual representation for participants choosing to begin with either a noun (animal) or an adjective (action).

For the two matching options in Step 1, selecting an adjective (e.g., "der weinende" [the crying]) would result in a pre-nominal description, whereas selecting the noun (e.g., "der Hase," [the rabbit,]) would lead to a post-nominal structure. The two distractors, which did not match the target properties, were also presented as options that could be selected to ensure participants could proceed to Step 2.

Step 2 implemented a linking rule that connected the participants' choice in Step 1 to the current visual scene. In Step 2, the buttons always listed all the properties that could be modified based on the participant's choice in Step 1, given the current trial display. This linking rule meant that Step 2 was conditioned by participants' choice made in Step 1, resulting in either two or five button options. Selecting the informative option in Step 1 narrowed the number of applicable figure candidates from 10 to 2, while the uninformative option narrowed it from 10 to 5, which are made visible in the number options in Step 2.

This design further emphasized the visual-numeric contrast between the figures, corresponding to RER, to help participants focus on the informativity aspect of a display. However, it is noteworthy that Step 2 might subtly encourage participants

to select the informative option in Step 1, as choosing between two options could require less cognitive effort than choosing between five. Despite this potential bias, the primary goal of the current experiment is to determine whether there is any effect of informativity on property ordering. If speakers' ordering strategies are still not affected by informativity in this experiment, even with this bias, it would provide strong evidence that informativity does not influence linearization at the syntactic level.

In the filler trials, Step 1 of the Maze Task also presented two matching options, structured similarly to those in the critical trials. Additionally, the two distractors in the critical trials were replaced with one-word options that can minimally specify the target (e.g., "der Hase?" [the rabbit?], "der Weinende?" [the crying?]), with the punctuation of a question mark signaling sentence ending. Choosing one of these options would directly complete the Maze Task. These minimal specification expressions differed qualitatively and syntactically from the full expressions in the critical trials, where selecting a singleton property could uniquely identify the target. The maze structure in the filler conditions introduced variations in expressions to help reduce the potential consistency effect across the critical trials (Tarenskeen et al., 2015).

4.1.1.5 Procedure

The experiment was implemented online via the LabVanced platform (Goeke et al., 2017). Before starting the experiment, a screen calibration procedure was conducted to ensure that the physical size of the visual displays was consistent across different screen devices used by the participants. In this procedure, participants were asked to drag and match a virtual card icon on the screen to the size of a standardized physical card (85.60 mm in width by 53.98 mm in height).

The experiment was introduced as a communication game in which participants collaborated with an online partner. The goal of the collaboration was to determine whether the displays shown to both the participants and their partners were identical, which was to be assessed based on the row location of the target figure. If the target appeared in the same row for both the participant and the online partner, the displays were considered identical; otherwise, they were not. In reality, the online partner was a simulated role, and their feedback was predefined by the experimenter.

The experiment began with a practice session to guide participants through the task. The first practice task was "responding", requiring participants to listen to their partner's descriptions (in reality, audio recordings) and to identify the location of the target. The second practice task was "asking", which was identical to the main experimental task, i.e., the Maze Task.

The detailed procedure for each trial in the main session is presented in Figure 4.3. Each trial started with a ten-figure display, with no indication of the target. After 2000ms, the target figure was highlighted with a red square. Participants were then guided to complete a question about the target's location, using the prompt "Wo ist _____?" [Where is _____?] , which appeared 2000ms after the

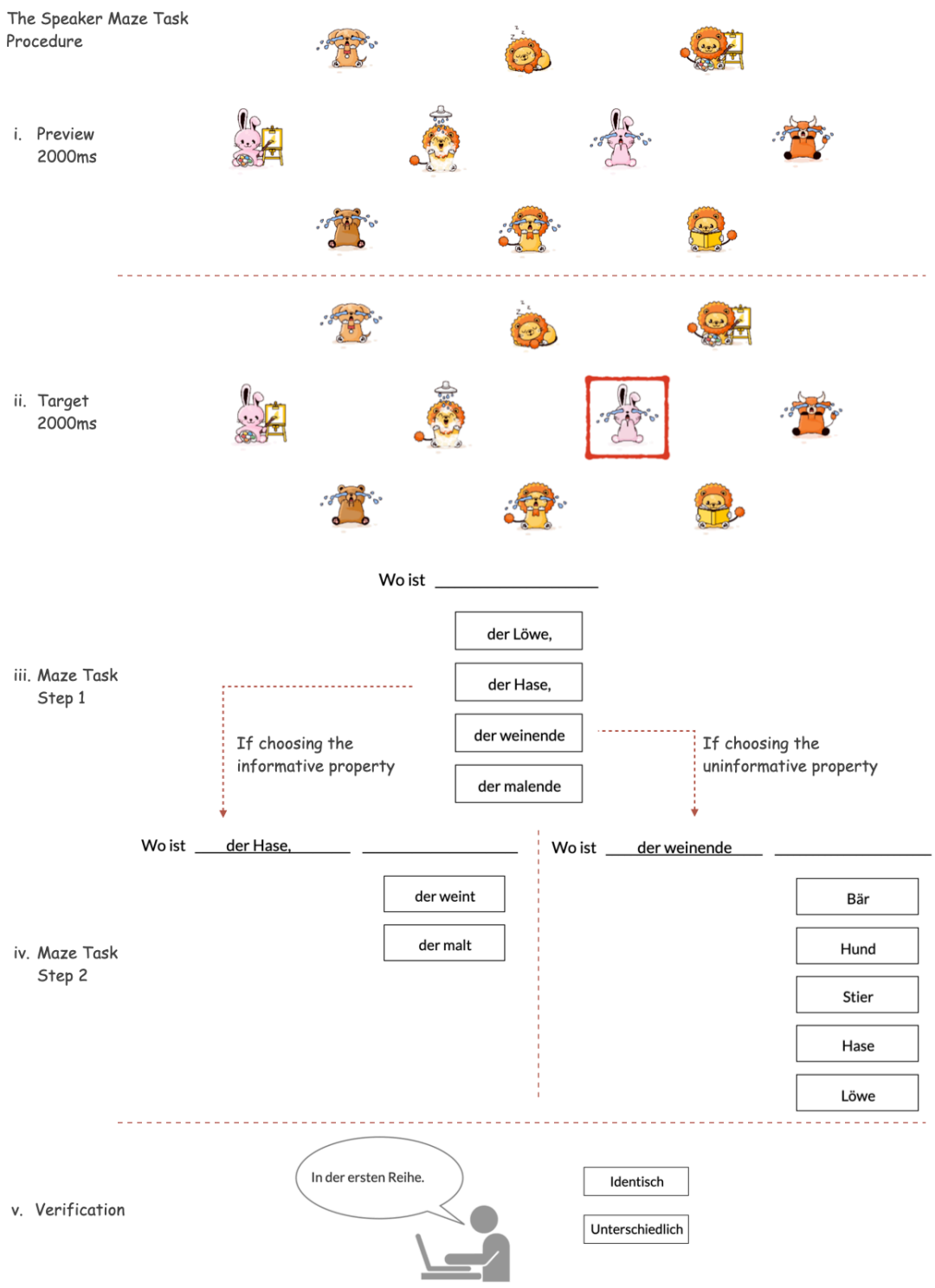


Figure 4.3: Experimental procedure of the Speaker Maze Task in Experiments 1-4: The display with the red square remained in the top half of the screen until the end of the trial, while participants completed the verification. The Maze Task and the verification procedure were presented in the lower half of the screen. This figure shows an example of the Animal-Informative Condition, where the participant and the partner viewed different displays.

target was highlighted. They completed the sentence by clicking through the maze (see Figure 4.2 and Figure 4.3).

Once Step 2 of the Maze Task was finished, the sentence was replaced by a textual reply from the partner, indicating the row in which the target was located from the partner's perspective. In reality, this reply was pre-defined, and there was a randomized waiting time of 1000-1500ms before the reply appeared, in order to simulate the time for processing by the partner. After a 500ms delay, two buttons ("same" or "different") were presented for verification. Participants clicked on one of these options to verify whether the displays were identical with their partner's, based on the partner's reply. One-third of the replies from the simulated partner suggested a mismatched display, where the target was found in a different row.

At the end of the experiment, participants completed a questionnaire to provide their demographic information and feedback on the experiment.

4.1.1.6 Analysis

A trial was coded as accurate when participants correctly described the target and accurately verified whether the displays were identical to their partner's. Participants were excluded if the proportion of inaccurate trials exceeded 10%. As a result, one participant was excluded from the analysis. Additionally, 40 inaccurate trials (1.05%) from other participants were excluded.

Participants' choices at Step 1 of the Maze Task were further coded into syntactic structure types: responses were classified as `pre-nominal` if the action property was selected at Step 1, or `post-nominal` if the animal property was selected. Proportional data and their associated confidence intervals were calculated in the same way as in Pre-test 1 (see Section 3.1.1.5).

A logistic mixed effect model analysis was conducted using R package `lme4` (Bates et al., 2015). The dependent variable was participants' choice at Step 1 in the critical trials, namely `Structure`: whether the description started with the action or the animal property, leading to one of two syntactic structures used for the descriptions: pre-nominal or post-nominal modifications, respectively. The predictor of the model was the `Condition` of the critical trials (the Action-Informative or the Animal-Informative Condition), which was dummy-coded, with the Action-Informative Condition as the baseline level.

Unless otherwise specified, the random effects structures of the statistical models reported in this thesis were selected based on the following criteria:

1) Successful convergence of a model,
2) When the difference of Akaike information criterion (AIC, Akaike, 1998) between two models was larger than 2, the model with the lower AIC would be selected.
3) If the difference of AIC was smaller than 2, the Bayesian information criterion (BIC, Schwarz, 1978) would be further compared. The model with the smallest BIC would be selected.

Both AIC and BIC are statistical tests that measure how well a model fits the data and thus serve as criteria for model selection by tradition. AIC was primarily considered because the current project sought to adopt the more fitted model with the maximal random effects structure (Barr et al., 2013), and AIC penalizes the more complex models to a lesser extent compared with BIC. If two model candidates are almost equally good, i.e., the difference of AIC is smaller than 2, then BIC comparison can further select the model that is better for the inference purpose (Ding et al., 2018).

Given that the variable `Condition` was the main predictor of interest for the entire thesis project, model comparisons came down to four random effects structures that differed in the inclusion or exclusion of the by-subject and/or by-item random slopes (listed below). For the current experiment, the random effects structure of the best model was `Model2`, including by-subject intercept, by-subject slope, and by-item intercept.

```
Model1:Structure~Condition+(1+Condition|subject)+(1+Condition|item)
Model2:Structure~Condition+(1+Condition|subject)+(1|item)
Model3:Structure~Condition+(1|subject)+(1+Condition|item)
Model4:Structure~Condition+(1|subject)+(1|item)
```

A power analysis for this experiment is reported in Appendix B, evaluating how the number of participants and critical trials in the current design relates to statistical power.

4.1.2 Results

4.1.2.1 Overall Results

Figure 4.4 presented the overall use of pre-/post- nominal structures in the two conditions. Regardless of conditions, there was a general preference for pre-nominal modifications, starting with describing the action property (78.47% pre-nominal vs. 21.53% post-nominal modifications, collapsed across conditions).

Among the post-nominal modifications across the two conditions (i.e., starting with the animal property), they were more frequently used in the Animal-Informative Condition (25.21%, proportion within condition), compared with the Action-Informative Condition (17.83%). This trend was, however, not significant ($\beta = 1.00, SE = 0.81, z = 1.24, p = 0.22$).

4.1.2.2 Group Results

According to Tourtouri et al. (2019), speakers' overspecification strategies give rise to group differences (see Chapter 2.1.2.2). Building on this finding, the current experiment identified group differences based on whether a participant exhibited any variability in their syntactic choices throughout the critical trials. This led to the categorization of participants into three groups:

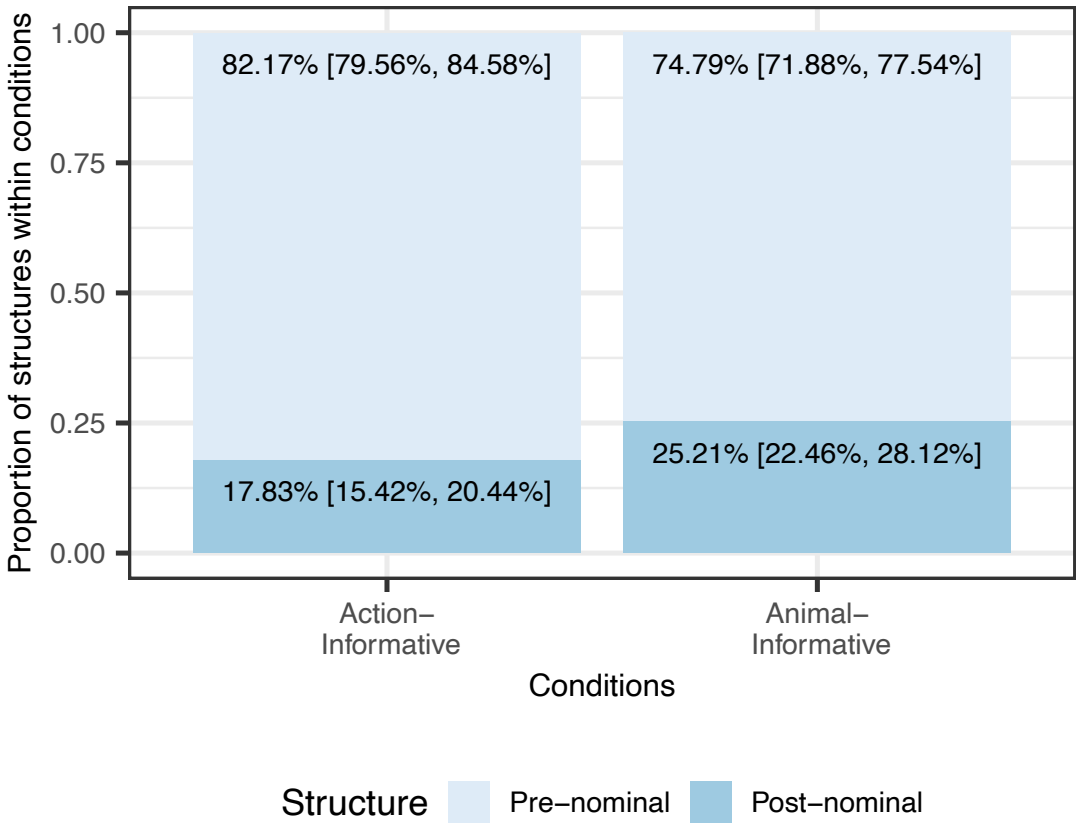


Figure 4.4: Experiment 1. Overall proportion of the use of pre-nominal (action-first) and post-nominal (animal-first) modifications within the two critical conditions. 95% CIs shown in brackets.

1. *Group Consistent.Pre-nominal*: 37 participants (46.84% of the 79 participants) always began with the *action* property, producing pre-nominal modifications for target descriptions.
2. *Group Consistent.Post-nominal*: 7 participants (8.86%) always began with the *animal* property, forming post-nominal modification structures.
3. *Group Varied*: The remaining 35 participants (44.30%) showed variability in their choices at Step 1, and so in the use of the two syntactic structures.

Figure 4.5 presents the distribution of the use of the two structures across these three groups.

A logistic mixed effect model analysis was conducted for *Group Varied* (N=35, the model was conducted in the same main effects structure as the model for the overall results, with the by-subject and by-item intercepts as the random effects structure), in which the frequency of the use of post-nominal modifications was significantly higher in the Animal-Informative than in the Action-Informative Condition ($\beta = 1.08, SE = 0.19, z = 5.76, p < .01$).

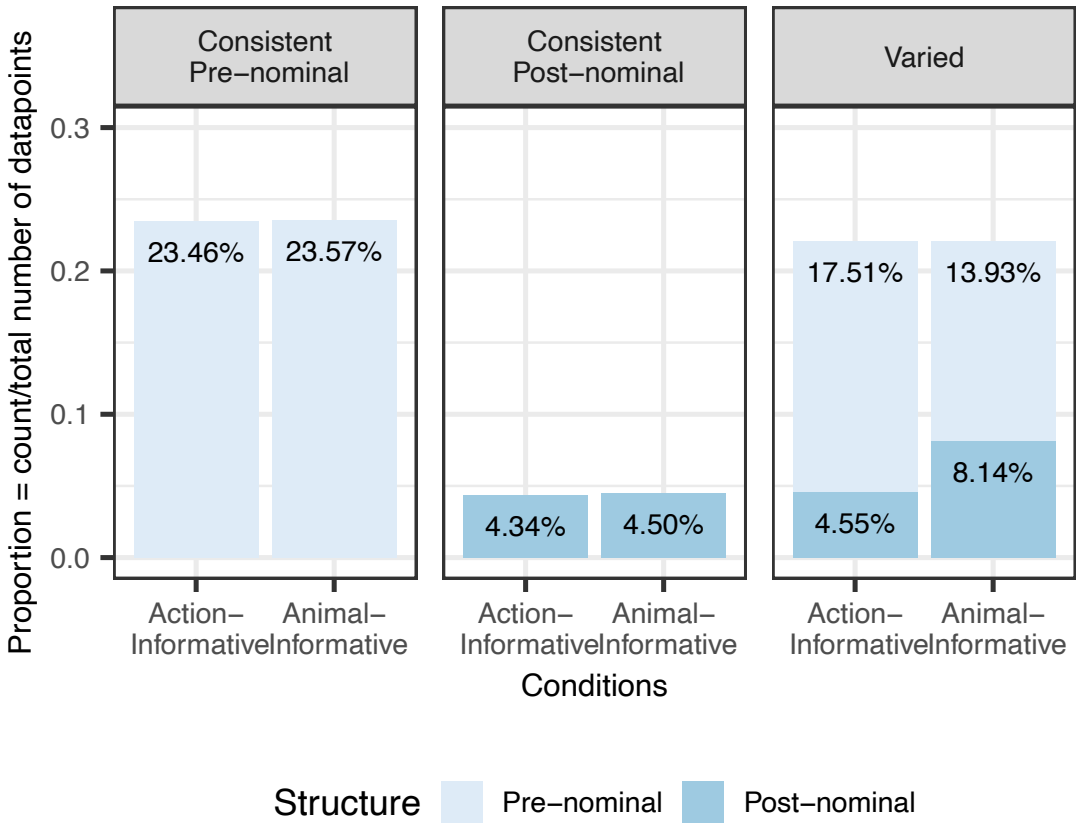


Figure 4.5: Experiment 1. Proportion of pre-nominal and post-nominal modifications within the three subject groups.

4.1.3 Discussion

In Experiment 1, participants engaged in an online referential communication game where they performed the role of speakers using a maze-based sentence completion task. The experiment aimed to compare the use of pre- and post-nominal structures in the Animal-Informative and Action-Informative Conditions, which differed in the informativity (i.e., RER) of the two properties. The more informative property reduced entropy to a greater extent (2.3 bits, narrowing the selection scope from 10 to 2 figures) compared to the less informative property (1 bit, narrowing the scope from 10 to 5 figures).

An overall preference for using the pre-nominal structure (beginning with the action property) was observed. When comparing the use of post-nominal structures (beginning with the animal property) across the two conditions, it was more frequently adopted in the Animal-Informative Condition. However, this trend did not reach statistical significance.

Tourtouri et al. (2019) observed group differences in speakers' overspecification strategies. One group of speakers consistently overspecified all properties of a target throughout the experiment, indicating a heuristic approach. In contrast, other groups overspecified more strategically based on RER, suggesting that these

speakers were sensitive to informativity and adopted a listener-oriented approach. Inspired by these findings, which suggest that both heuristic and listener-oriented strategies may coexist within population, the current experiment also identified group differences based on whether participants exhibited any syntactic variation throughout the experiment.

Group Consistent.Pre-nominal and *Group Consistent.Post-nominal* (*Group Consistent* as a whole) exhibited a fixed ordering of properties, either action-animal or animal-action, resulting in consistent use of a single syntactic structure (pre-nominal or post-nominal, respectively). For these groups, informativity did not influence property ordering at the syntactic level. Their ordering choices were instead driven by habitual preferences for one syntactic structure, unaffected by the information-theoretic properties of the visual context (i.e., RER).

In contrast, participants in *Group Varied* exhibited a more dynamic use of the two syntactic structures throughout the experiment. Comparing the two conditions, participants were more likely to select the informative property when both the informative and uninformative property options were available for selection in Step 1 of the Maze Task. This was particularly evident in the higher frequency of post-nominal modifications, as the less preferred order, in the Animal-Informative Condition compared to the Action-Informative Condition. These findings from *Group Varied* suggest that informativity influences property ordering at the syntactic level, increasing the likelihood that the more informative property is placed first in target descriptions, which supports an informative-first linearization preference.

Compared to the Pre-tests, Experiment 1 used ten-figure displays that were visually identical across conditions, achieving two key advancements. First, the contrast in RER between the action and animal properties was increased, enhancing the informativity contrast. The more informative property now would significantly facilitate target identification by narrowing down to a smaller selection scope (10 to 2 figures). Second, visual complexity, which was a confounding factor in the Pre-tests, was controlled across conditions, as the targets for both conditions were located within the same display.

In addition, the Maze Task provided a more sensitive method for capturing participants' linearization decisions by making both property orders, and thus the corresponding syntactic structures, explicitly available. By recording the first property participants selected, the task directly determined the order of the two properties and, consequently, the syntactic structure used in their target descriptions. This refinement enabled a more fine-grained assessment of how informativity influences property ordering at the syntactic level. Using the Maze Task, nearly half of the participants in Experiment 1 demonstrated variability in their syntactic choices throughout the experiment, marking a significant improvement over the dominant preference for pre-nominal structures observed in Pre-test 2. This was particularly notable given that the visual complexity in Experiment 1, with its ten-figure displays, may be even higher than the stimuli adopted in Pre-test 2 with only four figures, where a more heuristic approach might have been reinforced if using spoken production tasks.

However, more than half of the participants in Experiment 1 still preferred a consistent strategy, using only one of the modification structures for referential

descriptions in the Maze Task. The following two aspects of Experiment 1 may have contributed to this result, and the subsequent experiments aimed to address them:

First, the invariant behavior could be due to reduced engagement in the online experiment, where no co-present listener was involved. In Experiment 1, the simulated listener only provided textual feedback, which may diminish the sense of communication in the task. Participants might not have fully recognized that their descriptions were intended to guide the partner in locating the referent. Instead, they may have viewed the task as merely describing a target for themselves rather than for an audience.

Participants' sensitivity to informativity might improve if they first performed the role of the listener, receiving informative messages during target search. Previous research highlights the influence of role-switching in communication: for instance, Sikos et al. (2021a) found that pragmatic reasoning was enhanced when listeners first experienced the role of a speaker in an experiment. Vogels et al. (2020) observed that speakers' OS patterns became more listener-oriented after completing a challenging Listener Task. Building on this, Experiment 2 (Section 4.2) investigated whether performing a Listener Task first could amplify the impact of informativity on linearization choices in the subsequent Speaker Task.

Second, the invariant behavior may be influenced by participants' passive viewing of the visual displays before producing their descriptions. Participants may have focused solely on the target referent, disregarding the other nine figures in the visual context. This passive viewing strategy would make RER, revealed by the numerbic variations in the selection scope in the displays, irrelevant to their structural choices for referent descriptions. To address this, Experiment 3 (Section 4.3) expanded on Experiment 2 by animating the visual presentation of the ten-figure displays in the Speaker Maze Task. Figures were presented sequentially in different groups (e.g., first 2 rabbits, then 5 lions, and finally the remaining figures), forcing participants to process the entire display more thoroughly. Experiment 3 examined whether this adjustment could amplify participants' sensitivity to informativity in their linearization choices.

4.2 Experiment 2

To enhance participants' involvement in the referential communication game, Experiment 2 introduced a Listener Task in the first block. In this task, participants were required to locate the target based on auditory descriptions that mentioned the more informative property first. Following this, in the second block, participants performed the same Speaker Maze Task as in Experiment 1. Experiment 2 aimed to explore whether performing the Listener Task first could amplify the effect of informativity on property ordering in the subsequent Speaker Task.

If performing the Listener Task in the first experimental block enhances the effect of informativity, two predictions that are not mutually exclusive are raised. First, it might lead to a larger overall effect size of informativity, possibly reflected in a larger proportional difference in the use of post-nominal structures between the two critical conditions. Second, it could increase the proportion of *Group*

Varied speakers and decrease the size of *Group Consistent*, as exposing to both syntactic structures in the Listener Task may enhance participants' sensitivity to the two structures in the following Speaker Task, encouraging more flexibility in their encoding decisions.

4.2.1 Method

4.2.1.1 Participants

159 native German speakers (74 female and 85 male) were recruited via Prolific (mean age = 27.77). Participants were paid according to the German minimum wage (13.94 euro/hour).

4.2.1.2 Conditions, Materials, and Lists

In Experiment 2, the Listener Task and Speaker Task blocks each consisted of 24 trials to maintain a similar experimental length as in Experiment 1 (i.e., 48 critical trials in total). Within each block, there were 12 trials for the two critical conditions, with 6 trials per condition, and 12 trials for the four filler conditions (3 trials for each, see below). The trials for the Listener Task and the filler tials for both tasks were identical across lists.

Speaker Task For the critical conditions in the Speaker Maze Task, the materials in Experiment 2 were the same as in Experiment 1. The same 24 visual displays for the critical conditions were used, but half of the displays were presented per list, with the remaining half split across two additional lists.

The filler conditions were expanded to include four conditions in Experiment 2. In Experiment 1, the two Singleton filler conditions provided one-word, minimal specification options in Step 1 of the maze (e.g., "der Hase?" [the rabbit?] or "der Weinende?" [the Crying?]). The inclusion of the question mark indicated the end of the Maze Task and distinguished these options from the other two options that allowed continuation to Step 2. However, the results of fillers from Experiment 1 revealed that some participants quickly learned to select only the options with question marks (17 participants always chose these single-word options in at least one filler condition).

To avoid this strategy, while maintaining the same maze structure as in Experiment 1, Experiment 2 introduced two additional filler conditions: the Equal2- and Equal5-Filler Conditions. In these conditions, both the animal and action properties of the filler target reduced equal entropy. For example, in Figure 4.1, the Equal2 filler target was *the painting rabbit*, where both *painting* and *rabbit* reduced the selection scope from 10 to 2 figures. Similarly, the Equal5 filler target was *the crying lion* (reduced from 10 to 5).

The maze structure for the Equal filler conditions was identical as the Singleton fillers, but the targets in the Equal filler conditions cannot be described by the single-word options in the Speaker Maze Task, as they did not uniquely identify the target but forming underspecifications. Together with the two Singleton filler

conditions from Experiment 1, these four filler conditions (3 trials each per block) were randomized and placed before each critical trial in both blocks.

Listener Task For the Listener Task, 24 unique displays were created (12 for the critical conditions and 12 for the filler conditions), following the same design as in Experiment 1 (see Sections 4.1.1.2 and 4.1.1.3). The auditory descriptions for the Listener Task always followed an informative-first linearization preference. In the Action-Informative Condition, the pre-nominal modification was used, and in the Animal-Informative Condition, the post-nominal modification was used. The descriptions were synthesized speech recordings that were generated via the CereVoice TTS system's Alex voice (Version 3.2.0). Pauses, jittered around 1500ms, were inserted before the description of each property chunk to allow participants to process the content of speech. Each auditory description was encoded with a rising tone for the first property, signaling that the speech was continuing, and a falling tone for the second property to mark the end of the description.

For the Singleton filler conditions in the Listener Task, minimal specifications were used in the auditory recordings, ending with a falling tone to mark the end of the question (e.g., "Wo ist der Hase?" [Where is the rabbit?], "Wo ist der Weinende?" [Where is the crying?]). For the Equal filler conditions, half of the descriptions were encoded as pre-nominal modifications, and the other half as post-nominal modifications, with tonal encoding consistent with the critical conditions.

4.2.1.3 Procedure

The general setup and practice session of Experiment 2 were identical to those in Experiment 1. At the end of the practice session, participants were instructed to wait for assignment to an online partner. Participants were told that the two tasks performed during the practice session were randomly assigned to the participants and their partners, but in reality, the participants were always assigned to the Listener Task in the first block. The Listener Task and Speaker Task were presented consecutively as two blocks for each participant.

Listener Task Participants in the Listener Task were told to find the location of the target referent after listening to the descriptions of the target from the partner. Participants were told that the descriptions were from their partner who was performing the Maze Task, but the descriptions were further synthesized into speech to transmit to them.

Each trial in the Listener Task began with a ten-figure display, with no indication of the target. The audio description played after 4000ms. Participants were then required to select the corresponding row in which the target was located by clicking one of three button options, each representing the first, second, or third row.

Speaker Task The procedure of the Speaker Maze Task was identical to Experiment 1 (see Section 4.1.1.4).

4.2.1.4 Analysis

For the Listener Task, accuracy was measured by the correct identification of the row in which the target was located. Accuracy in the Speaker Task was measured in the same way as in Experiment 1. Subjects with more than 10% inaccurate trials in either task were excluded. 10 participants were excluded in this step. Another 85 individual inaccurate trials (4.83%) from other participants were also excluded.

Participants' choices at Step 1 of the Maze Task were coded as pre- or post-nominal syntactic structures, similar to Experiment 1. Proportional data and their associated confidence intervals were calculated in the same way as in Pre-test 1 (see Section 3.1.1.5).

A logistic mixed effect model analysis was conducted in the same method as in Experiment 1. The random effects structure of the best model for this experiment included by-subject and by-item random intercepts as well as by-subject and by-item random slopes.

Group categorization of participants, i.e., determining whether a participant belonged to *Group Varied* or *Group Consistent*, followed the same binary criterion established in Experiment 1 (see Section 4.1.2.2): whether a participant exhibited any syntactic variation in the critical trials across the experiment. *Group Consistent* was further divided into *Group Consistent.pre-nominal* and *Group Consistent.post-nominal* based on the consistent syntactic structure a partipant adopted in the experiment. One of the hypotheses for Experiment 2 was that the distribution of these two groups would differ from that observed in Experiment 1, specifically predicting an increase in the proportion of *Group Varied* participants due to the additional Listener Task increasing task involvement.

To test this hypothesis, the total number of participants categorized as *Group Varied* or *Group Consistent* was first counted in each experiment, based on the criterion described above. A two-sided Two Proportion Z-Test (referred to as "Two Proportion Z-Test" below) was then used to compare the group distributions across experiments. This test is a standardized statistical method for assessing whether there is a statistically significant difference between the proportions of a certain group within two independent samples (e.g., Webb, 2023). In the present context, the two independent samples correspond to the separate samples of participants from Experiment 1 and Experiment 2. The test was implemented in R using the `prop.test()` function and was used throughout the project to evaluate whether the proportion of *Group Varied* participants changed under different experimental manipulations intended to increase task involvement and communicative engagement.

Appendix B also reports a power analysis for this experiment.

4.2.2 Results

4.2.2.1 Overall Results

```
## boundary (singular) fit: see help('isSingular')
```

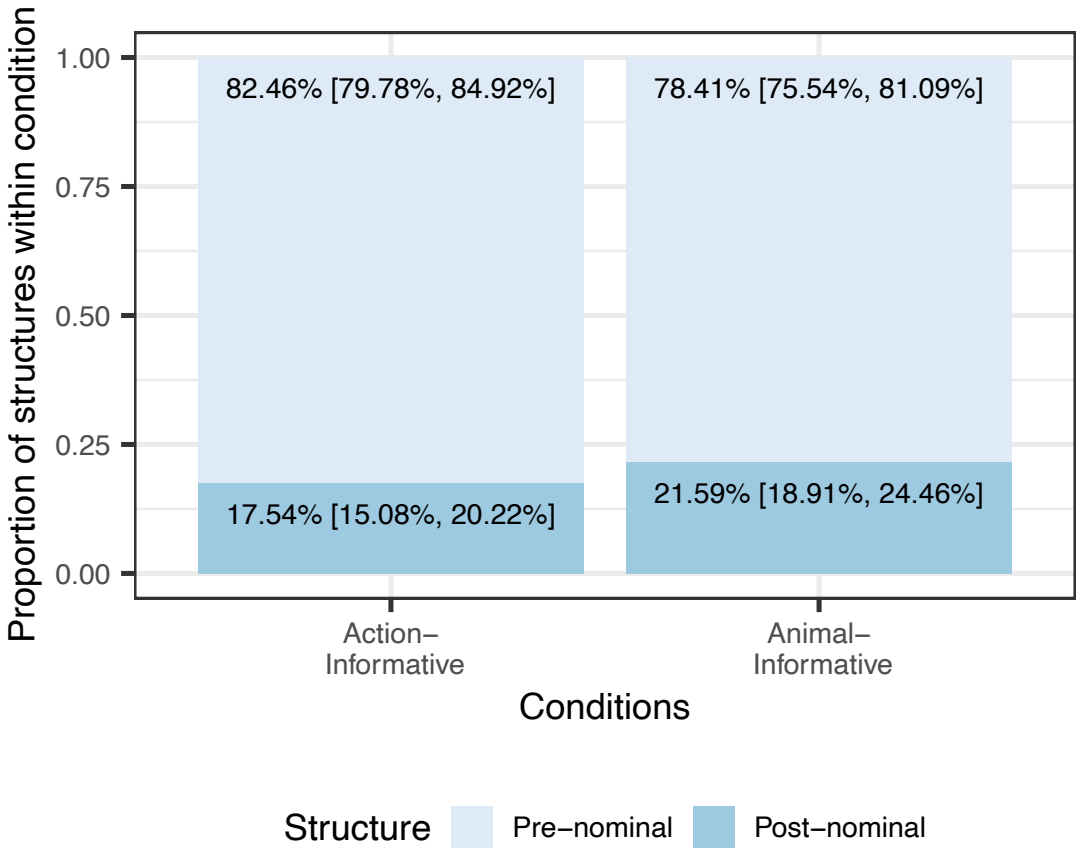


Figure 4.6: Experiment 2. Overall proportion of the use of pre-nominal (action-first) and post-nominal (animal-first) modifications within the two critical conditions. 95% CIs shown in brackets.

Figure 4.6 presented the overall use of pre-/post- nominal structures in the two conditions.

The result suggested that pre-nominal modifications were preferred overall (80.43% pre-nominal vs. 19.57% post-nominal modifications, collapsed across conditions). Among the post-nominal modifications across conditions, they were significantly more often used in the Animal-Informative Condition (21.59%, proportion within condition), compared with the Action-Informative Condition (17.54%) ($\beta = 1.44, SE = 0.52, z = 2.75, p < .01$).

4.2.2.2 Group Results

Similar to Experiment 1, the 149 participants can be categorized into three groups based on their syntactic consistency throughout the experiment: *Group Consistent Pre-nominal* obtained 74 subjects (49.66%); *Group Consistent Post-nominal* had 5 subjects (3.35%); 70 subjects (46.98%) belonged to *Group Varied*. Figure 4.7 presents the distribution of the three subject groups and their use of the two syntactic structures.

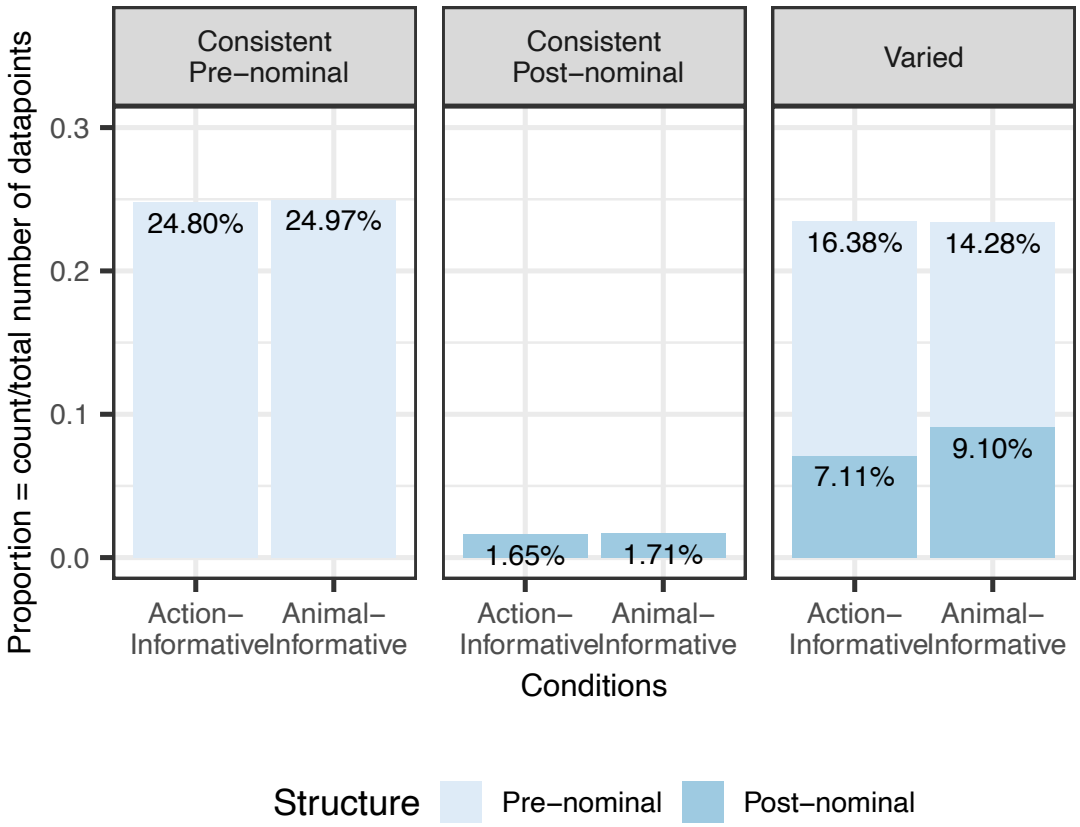


Figure 4.7: Experiment 2. Proportion of pre-nominal and post-nominal modifications within the three subject groups.

A logistic mixed effect model analysis was conducted for *Group Varied* (N=70, the random effects structure included the by-subject random slopes and random intercepts as well as the by-item random slopes and random intercepts), in which the frequency of the use of post-nominal modifications was significantly higher in the Animal-Informative Condition than in the Action-Informative Condition ($\beta = 0.68, SE = 0.29, z = 2.39, p < .05$).

The proportion of *Group Varied* participants in Experiment 2 (46.98%) did not significantly differ from that in Experiment 1 (44.30%), based on the Two Proportion Z-Test ($\chi^2(1, N = 228) = 0.59, p = 0.44$).

4.2.3 Discussion

Experiment 2 replicated the main findings of Experiment 1. Overall, pre-nominal modifications were preferred. At the same time, participants were categorized into three groups based on the variability of the use of the two syntactic structures. For *Group Varied*, their ordering pattern replicated the results from Experiment 1: post-nominal modifications were used more frequently in the Animal-Informative Condi-

tion compared to the Action-Informative Condition, consistent with an informative-first linearization preference.

However, the proportion of participants in *Group Varied* did not significantly increase compared to Experiment 1. The effect size of informativity remained small and comparable to that in Experiment 1, as reflected in the modest proportional difference (about 5%) in post-nominal modifications between the two conditions. Together, these results suggest that the Listener Task in the first block did not enhance participants' sensitivity to the informativity profile associated with different linearization choices in the Speaker Maze Task. This outcome implies that merely experiencing the role of a listener in a separate task block may not be sufficient to amplify the influence of informativity on speakers' linearization decisions.

4.3 Experiment 3

Experiment 3 aimed to encourage participants to attend more carefully to the entire visual context, thereby reducing the likelihood that they would fixate solely on the target and describe it in isolation. Experiment 3 was structured similarly to Experiment 2, but differed in how the displays were presented: in the Speaker Maze Task, the ten figures in the display were not presented simultaneously. Instead, the figures were presented consecutively in groups.

4.3.1 Method

4.3.1.1 Participants

100 native German speakers (32 female, mean age = 28.95) were recruited online via Prolific. They were paid according to the German minimum wage (13.94 euro/hour).

4.3.1.2 Conditions and Procedures

Conditions and Tasks were identical to Experiment 2 (see Section 4.2). The only difference was that in the Speaker Maze Task, the ten figures in the displays were presented in three groups, as presented in Figure 4.8: first, two figures with the same animal property appeared and stayed on the screen (e.g., in Figure 4.8, the two *rabbits* would first appear); after 1500ms, the other five figures with the same animal property were presented and stayed on the screen (the five *lions* in Figure 4.8); after another 1500ms, the rest three animals showed up and stayed on the screen for another 1500ms before target in the red square was highlighted. The presentation order of figures was identical for all conditions.

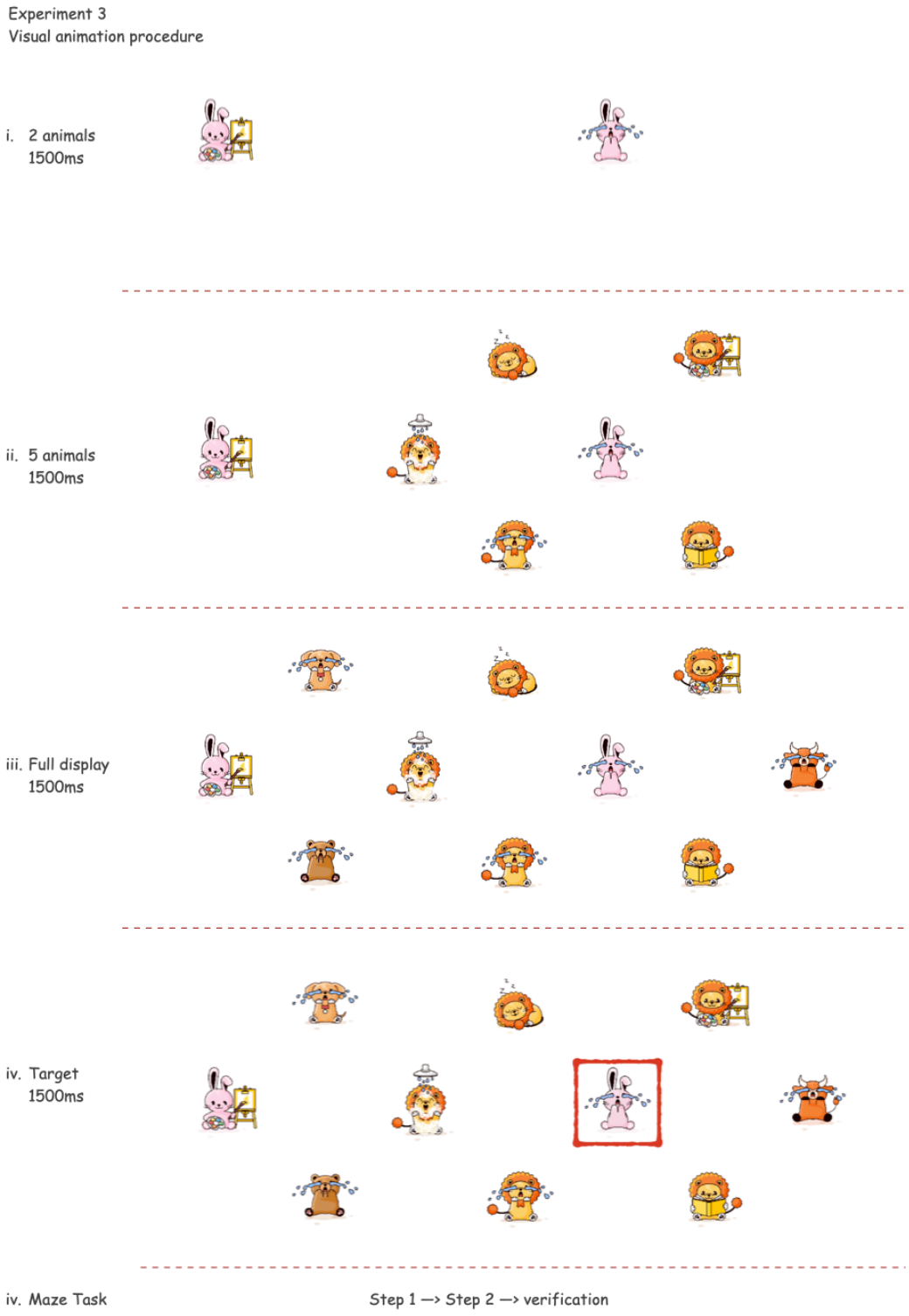


Figure 4.8: The sequence of figure presentations in the displays in Experiment 3. The procedure of the Maze Task was presented in Figure 4.3.

4.3.1.3 Analysis

Two participants were excluded based on the same criteria in Experiment 2. Another 62 individual inaccurate trials were further excluded.

Participants' choices at Step 1 of the Maze Task were coded as pre- or post-nominal syntactic structures, similar to Experiment 1. Proportional data and their associated confidence intervals were calculated in the same way as in Pre-test 1 (see Section 3.1.1.5).

The logistic mixed effect model analysis was conducted in the same way as in Experiment 1. The random effects structure of the best model included by-subject and by-item intercepts.

Two Proportion Z-Tests were conducted using the same method as in Experiment 2 to test whether there was a difference in the proportion of *Group Varied* participants between Experiment 3 and Experiment 2, as well as between Experiment 3 and Experiment 1.

4.3.2 Results

4.3.2.1 Overall Results

Figure 4.9 presented the overall use of pre-/post- nominal structures in the two conditions. Pre-nominal modifications were preferred overall (80.92% pre-nominal vs. 19.08% post-nominal modifications, collapsed across conditions). Among the post-nominal modifications across conditions, they were significantly more often used in the Animal-Informative Condition (21.48%, proportion within condition), compared with the Action-Informative Condition (16.67%) ($\beta = 0.69, SE = 0.23, z = 2.99, p < .01$).

4.3.2.2 Group Results

Similar to Experiments 1 and 2, participants were categorized into three groups: *Group Consistent Pre-nominal* obtained 53 subjects (54.08%); *Group Consistent Post-nominal* had 5 subjects (5.10%); 40 subjects (40.82%) belonged to *Group Varied*. Figure 4.10 presents the distribution of the three groups.

A logistic mixed effect model analysis was conducted for *Group Varied* (N=40, with the same random effects structure as in the previous model for 98 participants), in which the frequency of the use of post-nominal modifications was significantly higher in the Animal-Informative Condition than in the Action-Informative Condition ($\beta = 0.66, SE = 0.22, z = 2.92, p < .01$).

The proportional distribution of the three participant groups in the current experiment did not differ significantly from those observed in previous experiments. The Two Proportion Z-Tests showed that that the proportion of *Group Varied* participants in Experiment 3 (40.82%) was not significantly different from that in Experiment 1 (44.30%), $\chi^2(1, N = 177) = 0.22, p = 0.64$, nor from that in Experiment 2 (46.98%), $\chi^2(1, N = 247) = 0.91, p = 0.34$.

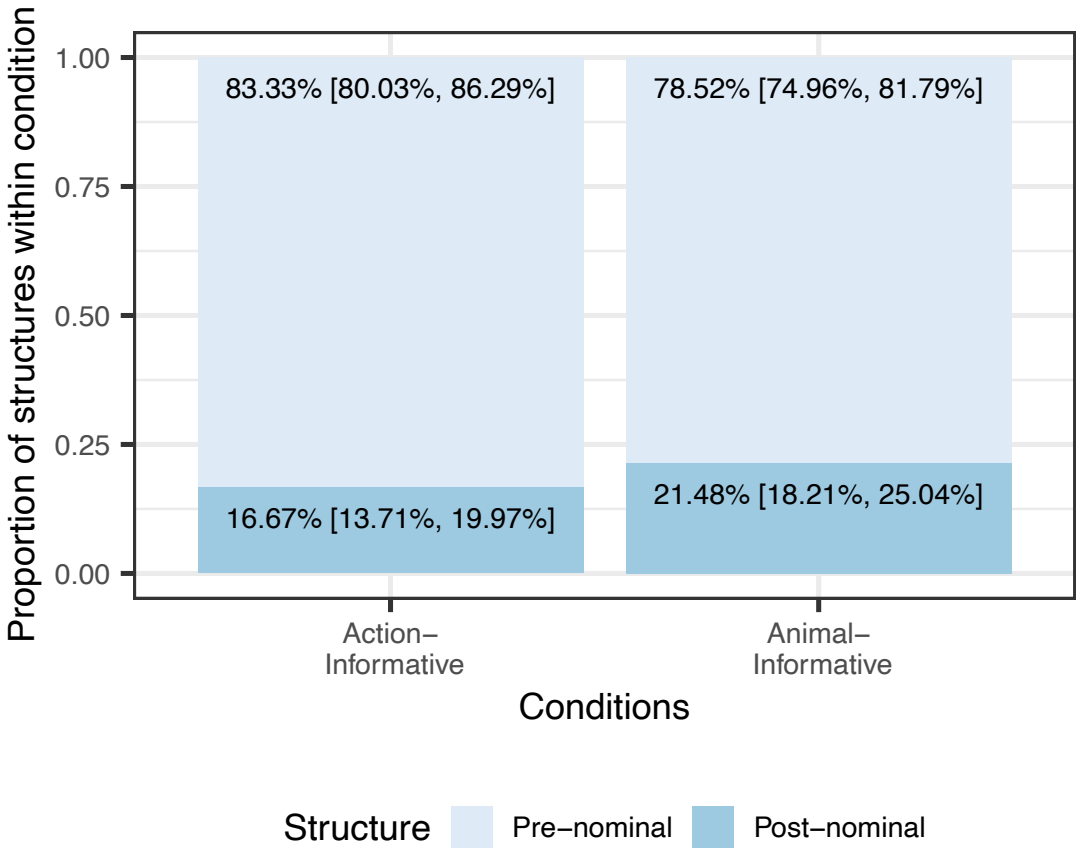


Figure 4.9: Experiment 3. Overall proportion of the use of pre-nominal (action-first) and post-nominal (animal-first) modifications within the two critical conditions. 95% CIs shown in brackets.

4.3.3 Discussion

Experiment 3 replicated the main findings of Experiment 1 and Experiment 2, with an overall preference towards the pre-nominal structure and a similar distribution of speaker groups where *Group Varied* was more likely to mention the more informative property first. Although not statistically significant, the proportion of *Group Varied* slightly decreased compared to Experiment 1 (40.82% vs. 44.30%) and Experiment 2 (46.84%). This result suggests that animating the ten figures to appear in three consecutive groups, intended to encourage more thorough comparisons between the target and the visual context, may have inadvertently distracted participants from the main task. Instead of enhancing their sensitivity to the informativity profiles of the linearization options, the animation might have introduced a higher cognitive load or disrupted participants' focus on the experimental manipulation.

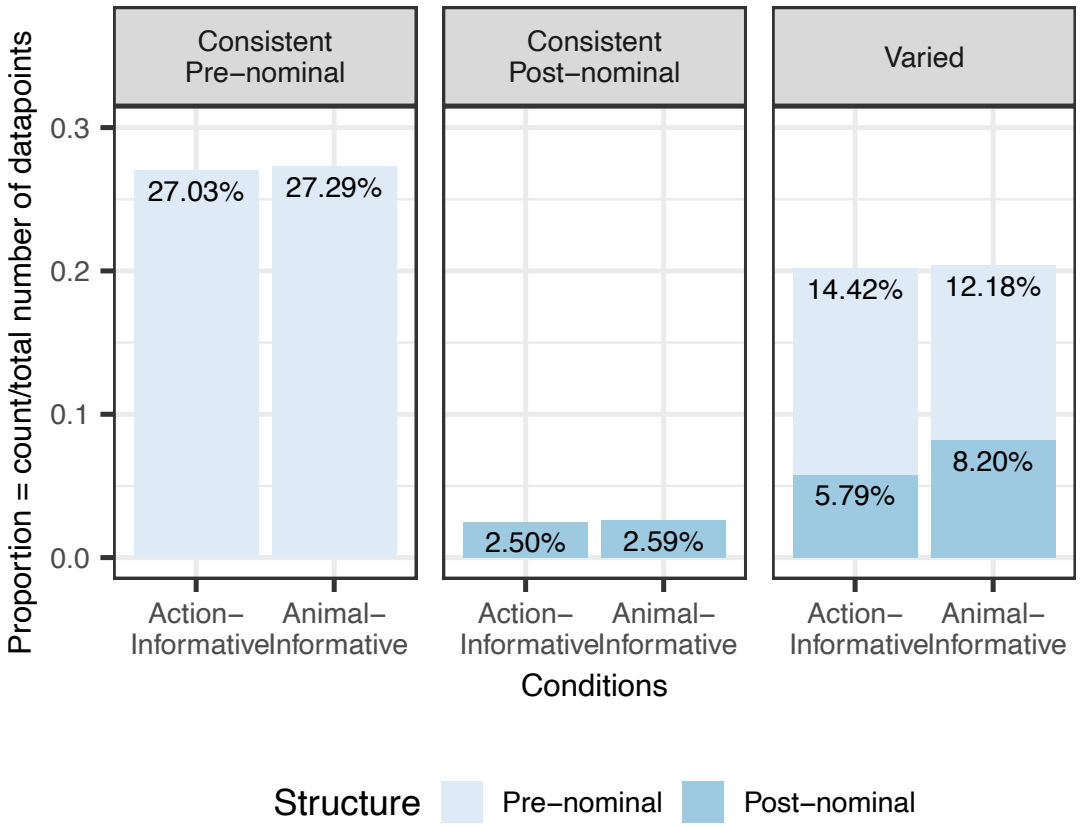


Figure 4.10: Experiment 3. Proportion of pre-nominal and post-nominal modifications within the three subject groups.

4.4 General Discussion

This chapter introduced three online referential communication experiments to examine whether informativity influences property ordering at the syntactic level and, if so, what the resulting ordering pattern might be. The experiments focused on participants' choices between pre-nominal and post-nominal modifications when describing visual scenes depicting animals performing actions. The visual displays presented ten-figure displays that differed in the selection scopes of the two properties, animal and action, creating distinct informativity profiles quantified by RER.

A Maze Task was employed to capture participants' syntactic choices by recording the very first property selected for referential expressions. This initial choice inherently determined the order of the two properties and the corresponding syntactic structure.

Experiments 2 and 3 introduced additional manipulations to increase participant involvement, aiming to amplify the effect of informativity beyond what was observed in Experiment 1. Experiment 2 included a Listener Task block to engage participants with the informative-first linearization descriptions for target search before performing the Speaker Maze Task. Experiment 3 further animated the

visual display, presenting figures in three consecutive groups to encourage more thorough comparisons between the target and the visual context.

To summarize the findings from the three experiments, pre-nominal modifications were generally preferred by German native speakers, consistent with the results of the Pre-tests. Across all three experiments, distinct group differences in speakers' linearization preferences were observed, which will be further discussed separately in the following sections: Approximately 55% of participants (*Group Consistent*) maintained a fixed property order throughout the experiment, predominantly using the action-animal sequence, which formed pre-nominal structures. Very few participants preferred the alternative animal-action sequence, which resulted in post-nominal modifications. The remaining participants (around 45%), categorized as *Group Varied*, demonstrated flexibility in their use of pre-nominal and post-nominal modifications, alternating the order of properties at least once during the experiment.

For *Group Varied*, informativity influenced linearization choices: comparing across the two conditions, the more informative property was mentioned first more frequently, that is, the post-nominal modifications (i.e., mentioning the animal first) were more frequent in the Animal-Informative Condition than in the Action-Informative Condition while the pre-nominal structures were more frequent in the Action-Informative Condition than in the Animal-Informative Condition. This modulation effect of informativity was statistically significant across all three experiments for *Group Varied*.

Experiments 2 and 3 successfully replicated the findings from Experiment 1. However, neither of the additional manipulations succeeded in amplifying the effect of informativity. Specifically, the frequency of post-nominal modifications did not increase further in the Animal-Informative Condition, nor did the proportion of participants in *Group Varied* increase. Although Experiments 2 and 3 also yielded statistically significant effects in the overall participant sample (i.e., including both *Group Varied* and *Group Consistent* speakers), unlike Experiment 1 where the effect was only significant among *Group Varied* speakers, this may primarily reflect the increased overall sample sizes in Experiments 2 and 3. Larger samples raise the likelihood of detecting statistically significant effects, even in the presence of relatively modest effect sizes (e.g., Andrade, 2020). A detailed examination of how participant and trial numbers influenced statistical power in these experiments is provided in Appendix B.

4.4.1 *Group Consistent*

The group distinctions observed across the three experiments suggested that speakers adopted different ordering strategies when formulating descriptions in referential communication tasks. For *Group Consistent*, the consistent ordering strategy reflected a speaker-oriented approach, where property ordering entailing syntactic variation was not influenced by informativity. This group consistently relied on a fixed order of properties, predominantly using the sequence action-animal, which formed the preferred, pre-nominal structures. Very few participants adopted

the alternative animal-action sequence, which resulted in post-nominal modifications. This consistent strategy aligns with the speaker-oriented, heuristic view for several reasons:

First, the consistent choice is likely driven by speakers' habitual preferences for specific syntactic structures, particularly pre-nominal modifications (action-animal). This preference may override the effect of informativity, as speakers are more likely to rely on habitual syntactic preferences when deciding on the structure for referential expressions rather than considering the listener's need for efficient target identification.

A similar trend was observed by Gatt et al. (2013) in property selection, where speakers' preference for certain modifiers (e.g., color > size > attachment) dominated over discriminability (see details in Chapter 2.1). Gatt et al. (2013) introduced the concept of "preference-based heuristics," suggesting that speakers may rely on a default preference for certain properties, shaped by their habitual preference towards mentioning a specific property or towards the use of specific linguistic structures. Similarly, the behavior of *Group Consistent* in the current experiments suggests that preference may also govern property ordering, with the pre-nominal structure, and thus the action-animal order, being preferred and more frequently adopted. This default decision may occur even before the first property is selected in the Maze Task, guiding speakers to map properties into their corresponding syntactic positions without considering other factors like informativity when formulating referential expressions.

Second, in addition to habitual syntactic preferences, the results in *Group Consistent* align with the consistency effect observed by Tarenskeen et al. (2015), as well as the syntactic priming effect (e.g., Hartsuiker, 1999; Jacobs et al., 2019). Tarenskeen et al. (2015) found that speakers who overspecify one property (e.g., size) were more likely to overspecify other properties (e.g., color), leading to a consistent tendency of adopting the same overspecification strategy within an individual speaker. Similarly, the syntactic priming effect refers to the tendency for speakers to repeat the syntactic structure used previously (see also Pickering & Garrod, 2004). Although the current experiments cannot fully disentangle the consistency effect from syntactic priming, since the same ordering strategy always leads to the same syntactic structure (i.e., the action-animal order always leads to pre-nominal structures), the pattern observed in *Group Consistent* suggests that these speakers are inclined to repeat the same referential strategy and linguistic structure across trials.

The effect of both a habitual syntactic preference and the priming effect (or the consistency effect) may further reinforce *Group Consistent* speakers to rely heavily on the preferred syntactic structure for linearization. This persistence is remarkable given the current experimental design, as it remained unaffected by the higher experimental task demands of the Speaker Maze Task adopted in this project. *Group Consistent* continued to select the same property in Step 1, regardless of the random positioning of the property options. They were also unaffected by the varied number of options in Step 2, even when choosing the uninformative property in Step 1 would require more effort to select from five options in Step 2. Moreover, Experiment 3, which introduced animated figure

presentation to encourage more thorough inspection of the visual displays, also failed to engage speakers in processing the numeric dimension of the displays in a way that influenced their property ordering choices.

The failure of these experimental attempts in changing the linearization strategy in *Group Consistent* highlights the robustness of the speaker-oriented heuristic approach, even under more complex tasks. This persistence calls for a deeper understanding of why speaker-oriented strategies are consistently preferred by some of the speakers. Chapter 6.2.3 will revisit this issue in reference production and provide some speculations.

4.4.2 *Group Varied*

The results from ***Group Varied*** suggest that informativity does play a role in influencing property ordering at the syntactic level. This finding directly addresses RQ1 and RQ2 of the project, suggesting that the modulation effect of informativity supports the informative-first linearization preference discussed in Chapter 1.4.1 and Chapter 2.2. Specifically, when comparing the two critical conditions, speakers in this group were more likely to place the more informative property earlier when ordering the two properties.

This preference aligns with findings from Fukumura (2018), who observed similar effects in color-pattern adjective ordering. The current results extend this effect of informativity to the syntactic level, where prioritizing the more informative property leads to alternations between pre-nominal and post-nominal modifications. These syntactic alternations are potentially more demanding than simply reordering adjectives, as producing post-nominal modifications requires qualitatively distinct and more effortful planning for production compared to pre-nominal modifications (Brown-Schmidt & Konopka, 2008, also discussed in Chapter 2.2.2.5). Despite this increased effort, speakers in *Group Varied* were still able to adjust their syntactic structures to prioritize the more informative property, probably motivated by a listener-oriented intention that the informative-first order could facilitate listeners' efficient identification of the target in their visual search (e.g., Rubio-Fernández et al., 2021). This pattern also highlights speakers' efforts to achieve communication efficiency. A more theoretical discussion of the informative-first linearization pattern and its connection to communication efficiency and the listener-oriented view will be revisited in Chapter 6.2.1.

The group distinction in speakers' encoding patterns aligns with findings by Tourtouri et al. (2019), who observed similar divisions in speakers' OS strategies, particularly in property selection. Tourtouri et al. (2019) identified one group of speakers who consistently overspecified all properties or only the color property, remaining insensitive to the RER of properties based on the visual contexts. In contrast, another group's OS patterns were systematically modulated by RER variations across conditions, demonstrating that OS can be modulated by informativity to facilitate listeners' target search. The current three experiments found a similar group distinction, but in terms of speakers' sensitivity to RER in property ordering. This group distinction extends from the OS strategies observed by Tourtouri et al. (2019) to property ordering, both suggesting the coexistence of the two types

of speakers preferring either a speaker-oriented or a listener-oriented strategy in reference production. This group division will be further discussed in Chapter 6.2.2.

4.4.3 Perspective Changing in Block Design

The additional experimental manipulations in Experiments 2 and 3, which introduced a Listener Task in the first experimental block before the Speaker Maze Task, did not enhance speakers' sensitivity to informativity in property ordering. Neither the effect of informativity nor the proportion of *Group Varied* participants increased relative to Experiment 1. This outcome may not be entirely surprising when considering the results of some previous studies.

Sikos et al. (2021a) and Vogels et al. (2020) also alternated participants' roles as speakers or listeners between blocks to investigate whether changing participants' perspectives between the two roles would change their performance as the speaker. However, their results indicated that alternating roles between blocks was either ineffective or had only moderate effects on improving participants' performance. Sikos et al. (2021a) evaluated human listeners' pragmatic reasoning in RSA-style referential games conducted online. They found that participants' pragmatic reasoning as listeners improved after performing as speakers first, as evidenced by higher accuracy in identifying targets after receiving one-word guidance. However, their performance as speakers remained unchanged, regardless of whether they had previously performed a Listener Task. Vogels et al. (2020) found that participants were more likely to overspecify as speakers to help guide listener-drivers. However, this trend only emerged after participants had experienced the more difficult listener-driving task as listeners. If the listener-driving task in the first block was easy, or if participants played the role of speaker in the first block without experiencing the listeners' perspective first, the OS patterns were similar. These evidence suggested that changing participants' roles between blocks may not be strong enough to influence speakers' referential strategies, unless the manipulation in the first Listener Task block was strong enough to make participants aware of the potential difficulties faced by the listeners (e.g., easy vs. difficult driving task, see details in Chapter 2.1.3.1).

The results of Experiments 2 and 3 in this chapter align with these findings, showing that performing a Listener Task first did not impact participants' sensitivity to informativity in the Speaker Task, at least not reflected by their sensitivity towards different property ordering encoded in different syntactic structures. This suggests that alternating roles across blocks, especially when the Listener Task block is passive or easy, may have only a subtle influence on participants' referential strategies as speakers in the later block. The influence of perspective-taking on speakers' referential strategies will be further discussed in Chapter 6.2.2.2.

4.4.4 Conclusion

Taken together, Experiments 1-3 showed that while *Group Consistent* adhered to a fixed syntactic structure and property order for reference production regardless of informativity, informativity indeed influenced property ordering in *Group Varied*, who were more likely to place the more informative property earlier in referential expressions when comparing the two conditions. However, the proportion of speakers in *Group Varied* and the effect size of informativity remained relatively small across all three experiments, despite additional manipulations in Experiments 2 and 3 aimed at increasing task involvement.

The next chapter, Chapter 5, addresses RQ3 of this project, examining whether and how communication dynamics might reinforce the effect of informativity observed in this chapter. Similar to the hypotheses for Experiments 2 and 3, this reinforcement effect may involve increasing the proportion of speakers in *Group Varied* and/or amplifying the effect size of informativity within this group. Experiment 4 increased communication engagement by alternating the Listener Task and the Speaker Task on a trial-by-trial basis, creating an interactive dynamic that requires constant change of perspectives. Experiment 5 further emphasized the sense of communication by conducting the referential communication game in a face-to-face communication environment, where speakers provided target descriptions orally to a confederate partner.

5

Informative Linearization in Interactive Communication

Contents

Experiments 1-3, introduced in the previous chapter, compared speakers' use of pre- and post-nominal modifications in describing target referents in the Animal-Informative and Action-Informative Conditions that varied the RER of the animal and action properties. These experiments demonstrated that, although with an overall preference of using the pre-nominal structure, *Group Varied* speakers were sensitive to informativity in syntactic linearization, preferring to use the syntactic order that places the more informative property earlier in referential expressions when comparing their syntactic choices between the two conditions. In contrast, *Group Consistent* speakers adhered to a consistent strategy, predominantly using the pre-nominal structure. Despite additional manipulations in Experiment 2 and

Experiment 3 encouraging task involvement, i.e., including a Listener Task block before the Speaker Maze Task and visual animation for display previewing, the distribution of *Group Varied* and the effect size of informativity remained unchanged.

Building on these findings, while Experiments 1-3 addressed whether and how informativity influences speakers linearization strategies, Experiments 4 and 5 focus on *where* and *when* these effects occur, specifically exploring the role of communication engagement in different referential communication environments, to address RQ3 of the current project: how different communication environments with different level of engagement influence speakers' linearization strategies.

The primary goal of these experiments was to test whether a more interactive communication environment would encourage speakers to adopt the informative-first linearization strategy, with the listener-oriented intention to facilitate listeners' target search. This hypothesis is grounded in the idea that as communication dynamics become more interactive, speakers are more likely to consider the needs of the listeners, and thus are more likely to be listener-oriented in their referential strategies. In this project, this would potentially manifest in their linearization patterns, i.e., ordering the properties in a way that facilitates the listener's target search incrementally (see Chapter 2.1.2.2 and Chapter 2.2.2.1).

To investigate this, Experiments 4 and 5, building upon the comparison between the Animal-Informative and Action-Informative Conditions established in Experiments 1-3, introduced two new manipulations to increase communication interaction and engagement of the experimental task. First, in both experiments, participants were required to alternate between the speaker and listener roles on a trial-by-trial basis. This design aimed to reinforce frequent changes of participants' perspectives as the listener and the speaker (see Chapter 2.2.2.4) and consequently, increase the likelihood of adopting listener-oriented strategies that may involve adopting the informative-first linearization preference.

Second, the communication environments of the two experiments were varied: Experiment 4 involved an online communication environment similar to Experiments 1-3, while Experiment 5 was conducted in a face-to-face lab setting, with participants seated across from each other and separated by a screen, allowing for in-person interaction while maintaining controlled visual access to the referential displays. The two communication environments allowed for a comparison of how different environments would influence communication engagement based on the type of addressee (a virtual or human addressee), with face-to-face communication being potentially more engaging than web-based communication (e.g., Cowan et al., 2023; Peña et al., 2023), which may further affect speakers' linearzation strategy in reference production.

The first manipulation, trial-by-trial role alternations, was implemented in both experiments, requiring participants to alternate between the Listener and Speaker Task on every trial to ensure frequent change of perspectives. However, this design would inevitably introduce another factor influencing property ordering, which is the priming effect, where speakers may tend to repeat the syntactic structure used by the Confederate in the preceding trial (e.g., Pickering & Garrod, 2004). A secondary question addressed in this chapter was how priming interacts with

informativity, which has been previously explored by Haywood et al. (2003) and Fukumura (2018) discussed in Chapter 2.2.2.3.

The second manipulation involved comparing the two communication environments: online communication versus face-to-face communication. Previous research suggests that speakers tend to exhibit more speaker-oriented behavior in online communication compared to face-to-face communication, as virtual addressees reduce engagement and authenticity (e.g., Cowan et al., 2015; Dombi et al., 2022; Peña et al., 2023). For instance, human speakers in *human-to-computer communication (HCC)*, compared with in *human-to-human communication (HHC)*, are more likely to provide privileged knowledge, such as using a phrase like "the *small* bone" when the contrastive object (e.g., "the big bone") is visible only to the human speaker but not to the computer addressee (Peña et al., 2023). This difference between HCC and HHC arises because the computer addressee reduces engagement and authenticity of communication. Experiment 4 (online) and Experiment 5 (face-to-face) tested whether different communication dynamics would influence speakers' sensitivity to informativity, specifically in property ordering, reflecting the extent to which speakers consider the type of addressee in their referential strategies.

Two alternative hypotheses are proposed when comparing Experiments 4 and 5, and when comparing them with Experiments 1-3 in the previous chapter: First, if more instances of informative-first linearization preferences are observed, and/or if a greater proportion of speakers are classified as *Group Varied*, especially in the face-to-face communication in Experiment 5 that is supposed to be the most engaging, this would suggest that perspective-changing encourages speakers to adopt listener-oriented strategies.

Alternatively, if the behavior of speakers in Experiments 4 and 5 remains consistent with that observed in Experiments 1–3, despite the two manipulations designed to increase communication engagement, it would suggest that perspective-taking and increased communication engagement do not significantly improve speakers' sensitivity to informativity, at least not as reflected in property ordering at the syntactic level.

Finally, a secondary hypothesis regarding the priming effect was proposed: speakers would tend to repeat the syntactic structure adopted by the Confederate in the preceding trial. The detailed effects of the interaction between priming and informativity in property ordering were left open for exploration. Previous work, such as Fukumura (2018) and Haywood et al. (2003), has shown that both priming and informativity can jointly affect property ordering, which has been discussed in Chapter 2.2.2.2.

5.1 Experiment 4

In Experiment 4, participants played an online referential communication game similar to Experiments 1-3 with an online simulated partner (referred to as the "Confederate"). Instead of the block design for the Listener and Speaker Maze Task used in Experiment 2, participants and the Confederate alternated between

the two tasks on a trial-by-trial basis in this experiment. In one experimental trial, the Confederate took on the role of the speaker, describing the target while the participant acted as the listener and searched for the target. In the next trial, they switched roles: the participant became the speaker, describing the target in the Maze Task while the Confederate responded. This trial-by-trial role alternation resulted in a switch between the listener and speaker perspectives for the participants.

The primary comparison in Experiment 4 was the effect of informativity, quantified by RER, on property ordering, specifically, the use of pre- and post-nominal structures in the Animal- and Action-Informative critical conditions, which were identical to Experiments 1-3. If the trial-by-trial role alternation successfully encouraged perspective-taking and increased communication engagement, it would potentially result in an enhanced preference towards the informative-first linearization.

A secondary comparison in Experiment 4 explored the combined effect of informativity and priming, with priming based on the syntactic structure adopted by the Confederate in the trial before the participant performed the Speaker Maze Task. Priming was included as a control variable to account for any potential influence on property ordering and was counterbalanced across four syntactic structures to ensure it did not confound the analysis of informativity (see Section 5.1.1.3 below). A potential hypothesis, based on Fukumura (2018) and Haywood et al. (2003), suggests that both informativity and priming could influence property ordering. While speakers tend to repeat the structure used by the Confederate due to priming and alignment between interlocutors (Pickering & Garrod, 2004), they would also be influenced by informativity, placing the more informative property earlier in the referential expression when comparing the two critical conditions. Fukumura (2018) did not detect a statistical interaction between these two effects, leaving this question open for further exploration in the current experiment.

The last hypothesis for Experiment 4 concerned the two speaker groups identified in Experiments 1-3: while they are still expected to be both present in the current experiment, the proportion of *Group Varied* speakers would increase compared to Experiments 1-3 due to increased communication engagement and the priming effect of the trial-by-trial role alternation design, both of which were expected to influence the distribution of the two speaker groups:

With higher communication engagement, as participants became more interactive with the Confederate in the communication game, the listener-oriented referential strategy, where speakers adjust property ordering to facilitate the listener's target search, was expected to be more frequently adopted. In contrast, the speaker-oriented strategy of consistently using one syntactic structure was expected to decrease, leading to a higher proportion of *Group Varied* speakers and a lower proportion of *Group Consistent* speakers.

Regarding the priming effect, previous studies have shown that syntactic priming can increase the likelihood of speakers repeating the less preferred structure (e.g., Cleland & Pickering, 2003; Haywood et al., 2003), i.e., the post-nominal modification for the current project. Based on the group classification criterion in the current project (see Chapter 4.1.2.2), participants would be classified as

Table 5.1: Experiment 4. Overview of task assignments for participants and the simulated Confederate

	Confederate	Participant
Speaker Task	Respond with target location (pre-defined)	Maze Task and verification
Listener Task	Maze Task and verification (pre-defined)	Respond with target location

Group Varied even if they used the post-nominal structure only once throughout the experiment. This would also contribute to a higher proportion of *Group Varied* speakers, as priming would increase the chance of using the post-nominal structure.

5.1.1 Method

5.1.1.1 Participants

80 native German speakers (36 female, mean age 28.61) were recruited online via Prolific, paid based on German minimum wage (13.94 euro/hour).

5.1.1.2 Procedure

Experiment 4 was conducted online as a communication game, following a similar setup and procedure to Experiments 1-3. The key difference was that, starting from the practice session, participants alternated with the Confederate in describing the target figure on a trial-by-trial basis. When it was **the Confederate's turn** to describe the target, the procedure for the participants was the same as the **Listener Task** from Experiment 2, where they listened to the target descriptions and responded by indicating the target's location (see Chapter 4.2.1.2). When it was **the Participant's turn** to describe the target, they followed the procedure from the **Speaker Maze Task** in Experiments 1-3 (see Chapter 4.1.1.4 and Figure 4.3). The roles and corresponding tasks for each turn are summarized in Table 5.1.

5.1.1.3 Conditions

The main measure of interest in Experiment 4 was participants' referential expressions recorded during the Speaker Maze Task, i.e., their behavior in the Participant Turn. Therefore, the experimental design of conditions and trial orders was centered on counterbalancing the two critical conditions, i.e., Animal-Informative and Action-Informative Conditions, which manipulated informativity in RER, as well as counterbalancing the primed syntactic structures introduced by the conditions and syntactic structures used in the previous trial (i.e., the Confederate Turn, where participants performed the Listener Task). Table 5.2 presents an overview of how the conditions in the Confederate and Participant Turns were combined. Experiment 4 crossed informativity (2 critical conditions) with priming (4 syntactic structures), resulting in a 2 x 4 design, elaborated below:

Table 5.2: Experiment 4. Overview of condition combinations for the Confederate and Participant Turns across Round Types

Round type	Conditions: Confederate Turns	Conditions: Participant Turns
Critical Round	Animal-Informative (Post-nominal)	Animal-Informative
	Action-Informative (Pre-nominal)	Animal-Informative
	Animal-Informative (Post-nominal)	Action-Informative
	Action-Informative (Pre-nominal)	Action-Informative
Mixed Round	Singleton-animal filler (Noun)	Animal-Informative
	Singleton-action filler (Adjective)	Animal-Informative
	Singleton-animal filler (Noun)	Action-Informative
	Singleton-action filler (Adjective)	Action-Informative
Filler Round	Equal fillers (pre-/post-nominal)	Singleton-animal filler
	Equal fillers (pre-/post-nominal)	Singleton-action filler

Conditions in the Speaker Task (Participant Turns) In Experiment 4, informativity had two levels: Animal-Informative and Action-Informative, identical to Experiments 1-3. The displays and the Maze Task used in the critical conditions in the Participant Turn were identical to those in Experiment 1, using 24 displays, each containing two targets shown separately in two lists. These lists followed the same counterbalanced design as in Experiment 1 (i.e., 12 trials for each critical condition within each list. See design details of the critical conditions in Chapter 4.1.1.3 and Chapter 4.1.1.4).

In addition to the two critical conditions, two Singleton Filler conditions were included in the Speaker Maze Task, identical to those used in Experiments 1-3 (i.e., Animal-Singleton and Action-Singleton Filler, see Figure 4.1 in Chapter 4). A total of 12 filler trials (6 trials for each Singleton Filler condition) were included in the Speaker Task. In total, the Speaker Task contained 36 trials with 24 trials of critical conditions and 12 trials of filler conditions.

Conditions in the Listener Task (Confederate Turns) The trial-by-trial role alternation design would inevitably introduce the priming effect, where participants' syntactic choices would be influenced by the syntactic structure adopted by the Confederate in the preceding trial. Similar to Experiments 2 and 3, the Confederate's target descriptions were designed by the experimenter and conveyed through synthesized speech recordings, always following the informative-first linearization preference, leading to the pre-nominal structure to be adopted in the Action-Informative Condition, the post-nominal modification in the Animal-Informative Condition, a single noun for the Animal-Singleton Filler, and a single adjective for the Action-Singleton Filler (see details in Chapter 4.2.1.2).

To introduce variability in the syntactic structures participants were exposed to during the Listener Task, Experiment 4 presented all four syntactic structures and the corresponding conditions (i.e., pre- or post-nominal, and single noun or

adjective) in the Confederate Turns before the Speaker Task trials in a counterbalanced manner, with 3 trials for each structure, subdivided from the 12 trials of each critical condition. The priming effect, consisting of four levels corresponding to the four syntactic structures, was included as a secondary control factor[1] to account for its potential influence on property ordering.

When participants performed a Singleton Filler trial in their turn in the Speaker Task, it was preceded by a filler condition in the Confederate Turn. In these cases, the Confederate would describe a target in the Equal-Filler conditions (i.e., Equal2- and Equal5-Filler Conditions, see details in Chapter 4.2.1.2). The descriptions were further counterbalanced by pre- or post-nominal structures, with 3 trials using each structure, subdivided from the 6 trials of each Singleton Filler condition in the Speaker Task.

For the Confederate Turn, 36 unique displays were created, consisting of 12 trials using pre-/post-nominal structures in the critical conditions, 12 trials using a single noun or adjective in the Singleton Filler conditions, and 12 trials in the Equal-filler conditions. The pre-defined targets were described by the corresponding structures in synthesized speech recordings. Visual and auditory materials were generated similarly as in Experiment 2 (see Chapter 4.2.1.2).

5.1.1.4 Trial Orders

To alternate the two tasks on a trial-by-trial basis, the combination of conditions for each trial was controlled and counterbalanced to ensure that the critical conditions in the Participant Turn were equally affected by the conditions and syntactic structures used in the preceding Confederate Turn. Experiment 4 defined an experimental Round as a combination of a trial in the Confederate Turn followed by a trial in the Participant Turn. In total, 3 types of Rounds were included in Experiment 4:

1. Filler Round: Both trials presented the filler conditions.
2. Mixed Round: The Confederate Turn presented a filler condition, followed by the Participant Turn in a critical condition.

[1]The priming effect of the four syntactic structures was considered a secondary control variable, rather than a primary factor, for the following reasons: 1) Priming was an inevitable effect arising from the trial-by-trial role alternation design, where participants' syntactic choices are influenced by the structures adopted by their partner (Pickering & Garrod, 2004). However, the trial-by-trial design, in the sense of increasing communication engagement, was of better interest than the priming effect itself, thus priming was not the primary focus of the current project. 2) The total number of trials for each priming structure per critical condition was relatively small (3 trials), which could limit the statistical power to detect the effects of priming. 3) The conditions under which each syntactic structure was used were pre-defined based on the informative-first linearization preference (e.g., pre-nominal modifications were used only in the Action-Informative condition), allowing participants to experience the listener's perspective and benefit from the informative-first order for efficient target search. Counterbalancing the priming structures was incorporated secondarily to control for any confounding effects from the Confederate Turn. If purely considering priming as a primary factor, it would require priming to be independent of informativity and conditions so that any effect could be attributed solely to the syntactic structure.

3. Critical Round: Both trials presented the critical conditions.

Table 5.2 summarizes the detailed combinations of conditions within each round type. In the Critical Rounds, the two critical conditions in the Participant Turn were equally preceded by the two critical conditions in the Confederate Turn, as well as by the two Singleton-filler conditions in the Confederate Turn in the Mixed Rounds. This combination also ensured that participants were equally primed by pre- and post-nominal modifications, as well as by minimal specifications (single noun or adjective expressions). In the filler Rounds, the two Singleton Fillers in the Participant Turn were equally preceded by the two syntactic structures and the two Equal filler conditions in the Confederate Turn.

The three types of Rounds (two trials per round, one for the Confederate Turn and one for the Participant Turn) were grouped into a group, consisting of six trials. In each group, the Filler Round was always presented first to reduce the consistent tendency of production, and the order of the remaining two rounds was randomized. The groups, each consisting of three rounds, were then randomized into a full experimental list.

5.1.1.5 Lists

Experiment 4 contained two experimental lists. Each list included 36 trials for the Participant Turn (24 critical trials and 12 filler trials) and 36 trials for the Confederate Turn (12 critical trials and 24 filler trials). The critical conditions in the Participant Turns used the same 24 displays created in Experiment 1, following a Latin-Square design across the two lists. The stimuli used in the Confederate Turns and the fillers remained the same across the two lists, with the constraint that there was no property overlap between the target properties presented in the Confederate Turns and the target properties in the subsequent Participant Turn to avoid lexical repetition.

5.1.1.6 Analysis

The exclusion criteria were the same as in Experiment 2. One participant was excluded due to chance-level accuracy in the Maze Task, leaving a total of 79 participants for analysis. Additionally, 94 individual inaccurate trials were excluded.

Participants' choices at Step 1 of the Maze Task were coded as pre- or post-nominal syntactic structures, following the same procedure from Experiment 1. Proportional data and their associated confidence intervals were calculated in the same manner as in Pre-test 1 (see Section 3.1.1.5).

Participants were categorized into *Group Varied* or *Group Consistent* based on the same binary criterion established in Experiment 1 (see Section 4.1.2.2). To test whether the proportion of *Group Varied* participants in Experiment 4 differed significantly from that in Experiments 1-3, three Two Proportion Z-Tests were conducted to compare the proportion of *Group Varied* in Experiment 4 with each previous experiment, using the same method as in Experiment 2 (see Chapter 4.2).

To investigate the main effect of informativity on participants' syntactic choices in the Speaker Maze Task, the data were first analyzed using logistic mixed effects

modeling with `Condition` as the primary factor, following the model selection procedure used in Experiment 1 (see Chapter 4.1). The best model included random effects for by-subject intercepts, by-subject slopes, and by-item intercepts.

Following this, the analysis then tested the priming effects based on the syntactic structures used in the Confederate Turns on the Speaker Maze Task. Another logistic mixed effects model analysis was conducted, which included `priming` as a secondary control variable. The second model included the main effects of `Condition` (the Animal-Informative and Action-Informative Conditions) and `Priming` (the four structures used in the Confederate Turn: pre-nominal or post-nominal structures in the critical trials and a single noun or adjective minimal specification in the filler trials). The best model included random effects of by-subject and by-item random intercepts, as well as a by-subject random slope. `Condition` was sum-to-zero coded with two levels, and `Priming` was dummy coded with four levels, with the pre-nominal structure as the baseline level. This contrast coding scheme allows the model to compare each of the three priming structures to the baseline, i.e., the pre-nominal structure, while still including the main effects of `Condition`.

In addition to choices of syntactic structures, participants' reaction times (RTs) for selecting the first property in Step 1 of the Maze Task were also recorded. An exploratory analysis of these RTs is reported in Appendix A.1.

5.1.2 Results

5.1.2.1 Overall Results

Figure 5.1 presented the overall use of pre-/post- nominal structures in the two conditions. In general, pre-nominal modifications were preferred (82.83% pre-nominal vs. 17.17% post-nominal modifications, collapsed across conditions). Post-nominal modifications were more frequently used in the Animal-Informative Condition (19.87%, proportion within condition), compared with the Action-Informative Condition (14.47%). However, the difference was not significant ($\beta = 0.48, SE = 0.32, z = 1.50, p = 0.13$).

5.1.2.2 Group Results

The 79 participants can be categorized into two groups based on their syntactic consistency throughout the experiment: *Group Consistent Pre-nominal* obtained 28 subjects (35.44%); 51 subjects (64.56%) belonged to *Group Varied*. Figure 5.2 presents the distribution of structural use from the two groups.

A logistic mixed effect model analysis was conducted for *Group Varied* (N=51, model structure was the same as the previous model for 79 participants), in which the frequency of the use of post-nominal modifications was significantly higher in the Animal-Informative Condition (30.90% within condition in *Group Varied*) than in the Action-Informative Condition (22.35%) ($\beta = 0.56, SE = 0.21, z = 2.68, p < .05$).

The distribution of *Group Consistent* and *Group Varied* participants in Experiment 4 differed significantly from the distributions observed in the previous

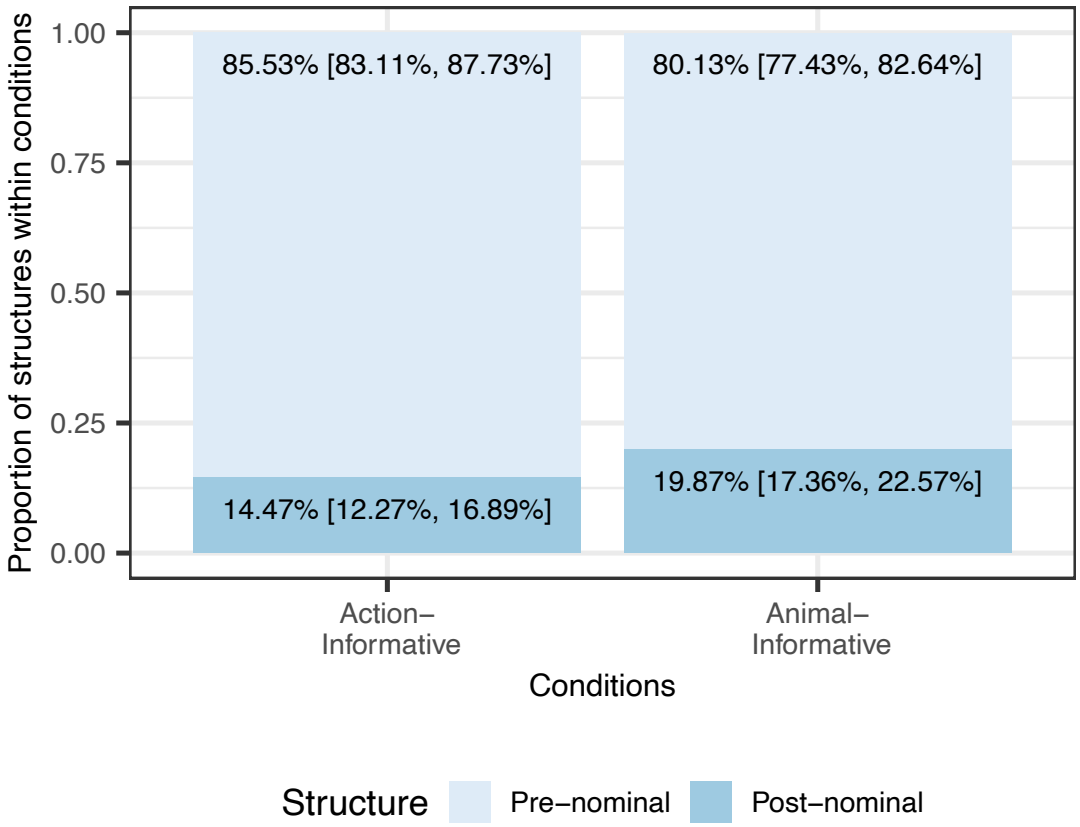


Figure 5.1: Experiment 4. the proportion of the use of pre-nominal (action-first) and post-nominal (animal-first) modifications within the two critical conditions. 95% CIs shown in brackets.

experiments. Two Proportion Z-Tests showed that the proportion of *Group Varied* participants in Experiment 4 (64.56%) was significantly higher than in Experiment 1 (44.30%, $\chi^2(1, N = 158) = 6.53, p < .05$), Experiment 2 (46.98%, $\chi^2(1, N = 228) = 6.40, p < .05$), and Experiment 3 (40.82%, $\chi^2(1, N = 177) = 9.87, p < .01$). These results suggest that the proportion of *Group Varied* participants in the current experiment was significantly greater than in all previous experiments.

5.1.2.3 Results of Priming and Informativity

The analysis below focused on the data from *Group Varied* (N=51), as *Group Consistent* was not significantly affected by either informativity or priming. Figure 5.3 depicts the proportion of post-nominal modifications, calculated as the number of post-nominal structures in each combination of a critical condition and a preceding Confederate structure, divided by the total number of trials for each combination. Table 5.3 summarizes the statistical results.

The analysis revealed a significant main effect of `Condition`, as well as a significant difference between the use of post-nominal and pre-nominal structures.

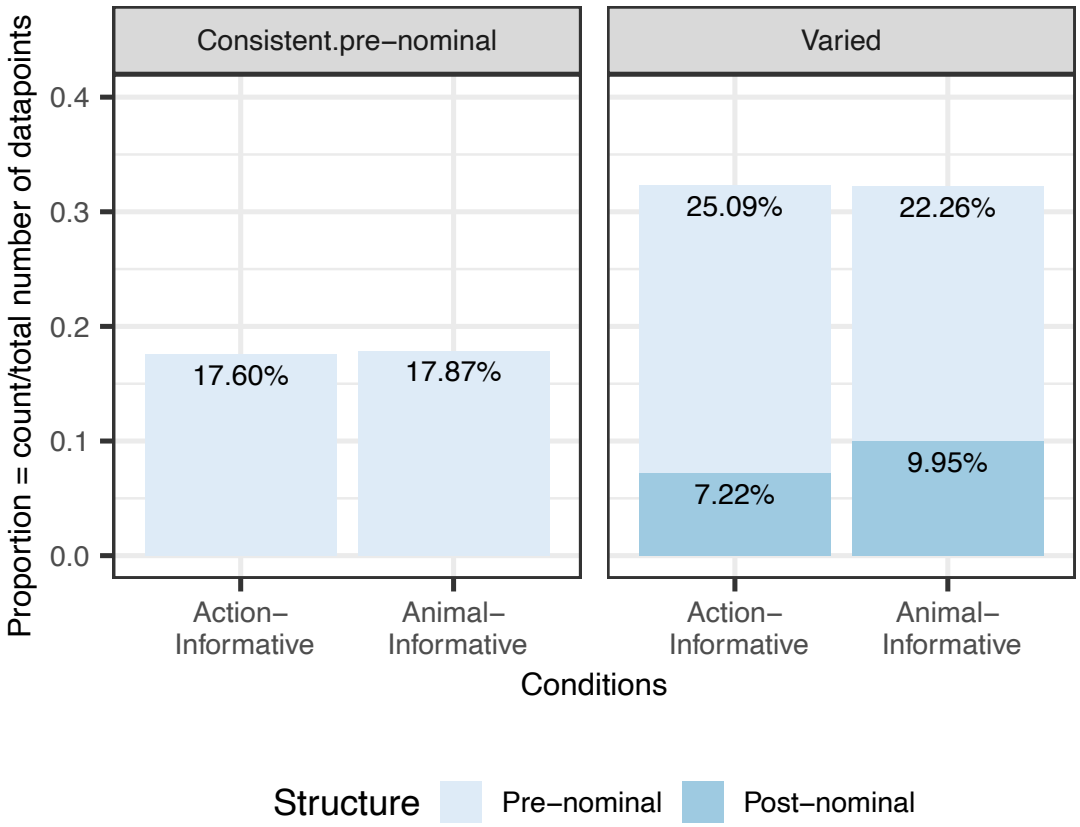


Figure 5.2: Experiment 4. Proportion of pre-nominal and post-nominal modifications within the two subject groups.

Table 5.3: Experiment 4. Results from the logistic mixed-effects model testing the main effects of Condition and Priming.

Predictor	$\hat{\beta}$	95% CI	z	p
Intercept	-1.69	[-2.33, -1.05]	-5.16	$< .001$
Condition: Animal vs. Action	-0.28	[-0.49, -0.08]	-2.70	.007
Priming: Post-nominal vs. Pre-nominal	0.97	[0.18, 1.75]	2.42	.016
Priming: Adjective vs. Pre-nominal	0.48	[-0.31, 1.27]	1.20	.230
Priming: Noun vs. Pre-nominal	-0.13	[-0.93, 0.67]	-0.32	.748

Note. The model was specified as: Structure ~ Condition + Priming + (1+ Condition|subject) + (1|item). Condition (Animal-Informative and Action-Informative) was sum-to-zero coded. Priming (pre-nominal, post-nominal, noun and adjective) was dummy-coded with the pre-nominal structure as the reference level. Each priming coefficient represents a comparison to the pre-nominal baseline.

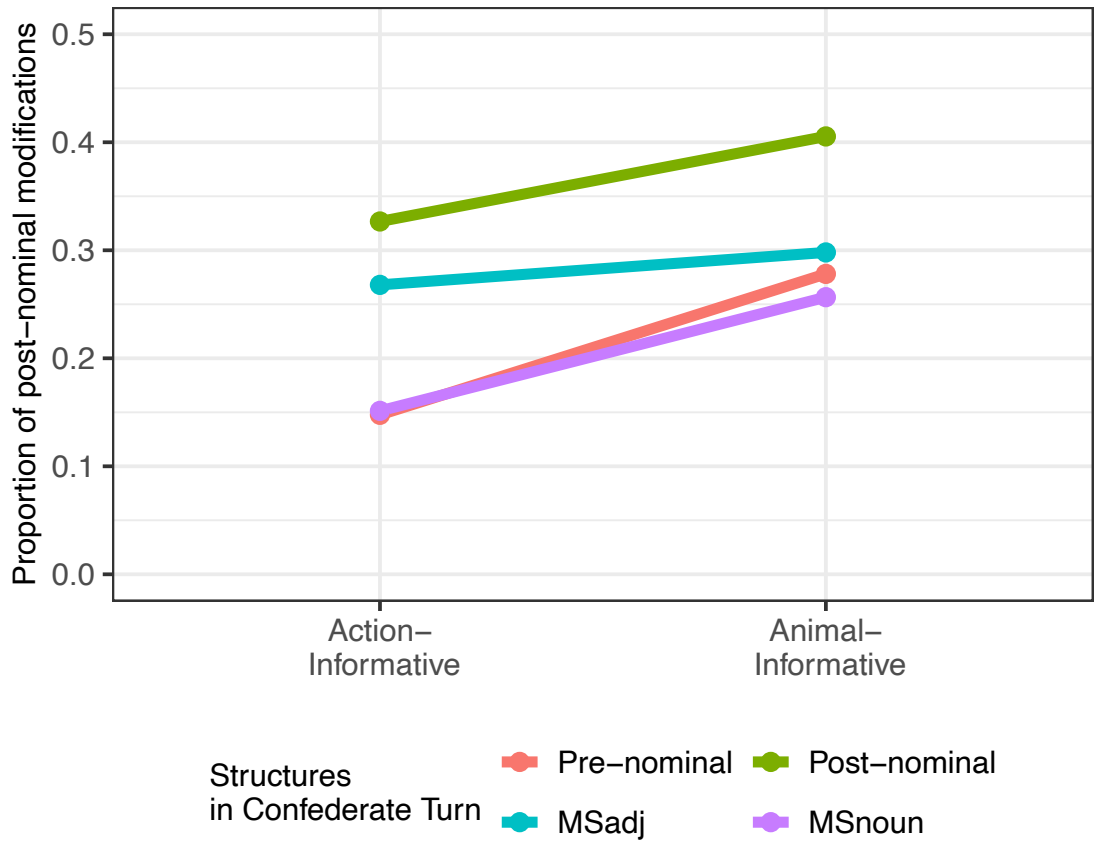


Figure 5.3: Experiment 4. Proportion of post-nominal modifications in each critical condition, grouped by the preceding syntactic structure used by the Confederate, for participants in *Group Varied*.

Specifically, participants primed by a post-nominal structure in the previous Confederate Turn used the post-nominal structure in their own turn more frequently. However, when participants were primed by a single noun or adjective minimal specification, no significant differences were observed.

An additional model that included the interaction between the two main effects was also tested. However, this model did not significantly improve model fit over the model reported in Table 5.3 ($BIC_{main} = 1297.28$, $BIC_{interaction} = 1309.37$, $\chi^2(1) = 0.00, p = 1.00$).[2]

5.1.3 Discussion

Experiment 4 built on the online communication environment established in Experiments 1–3 by introducing a simulated Confederate who alternated the speaker-listener roles with participants on a trial-by-trial basis. This role alternation was

[2]Given the smaller number of trials per priming condition and the limited number of trials involving post-nominal structures, models with the priming effect should be interpreted with caution due to low statistical power.

designed to encourage perspective-changing and communication engagement. The experiment aimed to investigate whether this manipulation would enhance the use of the informative-first linearization preference as a listener-oriented encoding strategy in property ordering.

The main findings replicated the effects observed in Experiments 1-3 with the same group divisions of speakers. Specifically, speakers in *Group Varied* exhibited a flexible use of pre-nominal and post-nominal modifications based on the informativity of the property words. In line with previous experiments, the informative-first linearization preference was observed: When comparing the two conditions, post-nominal modifications (starting with the animal property) were more frequently used in the Animal-Informative Condition, while the Action-Informative Condition showed a more frequent use of pre-nominal structures. In contrast, *Group Consistent* speakers adhered to the pre-nominal structure throughout the experiment.

Although the effect size of the modulation by informativity remained similar to previous experiments (with about a 5% increase in post-nominal structures in the Animal-Informative Condition), the proportion of *Group Varied* speakers significantly increased in Experiment 4 (65%) compared to Experiments 1–3 (less than 50%). More speakers were able to vary their use of the two syntactic structures forming the tendency to prioritize the more informative property early in their descriptions, compared to the previous experiments.

Regarding the priming effects, the syntactic structures used in the Confederate Turns were equally distributed before the critical conditions were presented to participants. Experiment 4 showed that the effect of informativity operated independently of the priming effect, particularly in Critical Rounds, where both participants and the Confederate encountered the critical conditions. While the participants did show a tendency to repeat the syntactic structure used by the Confederate in the previous trial, informativity remained a significant factor in guiding property ordering during the critical conditions.

The larger proportion of *Group Varied* observed in Experiment 4 can be explained by two possibilities. First, it suggests that the distribution of the two groups, *Group Varied* and *Group Consistent*, which has been previously identified in Experiments 1-3 and by Tourtouri et al. (2019), is not static. Instead, speakers' choice between being speaker- or listener-oriented appears to be significantly influenced by communicative factors, particularly the extent to which speakers are required to adopt the listener's perspective. In Experiment 4, the trial-by-trial role alternation may have emphasized the listener's perspective for participants, potentially encouraging more participants to vary their use of the two syntactic structures. This could further lead to a tendency to adopt the informative-first linearization strategy in property ordering, possibly reflecting an effort to facilitate the listener's target search more efficiently, informed by the speakers' immediate experience in the listener role.

Second, it is also possible that the proportional increase in *Group Varied* is due to the priming effect, where speakers are more likely to reuse the structures adopted by their partners. According to the criteria for group classification in the current project, participants would indeed be more likely to be categorized into

Group Varied if they were affected purely by syntactic priming, even if they used the post-nominal structure in just a single trial.

While these two explanations will be further examined in Experiment 5, the results from the Critical Rounds in Experiment 4 showed that informativity and priming independently influenced property ordering, with both factors playing significant roles and without a robust statistical interaction. Similar patterns have been observed in previous research on adjective ordering by Haywood et al. (2003) and Fukumura (2018), which will be further discussed in Chapter 6.

During the Mixed Rounds, where the Confederate produced a single noun or adjective in the Singleton-Filler conditions while participants performed the critical conditions, the priming effect appeared to be reduced, resembling the baseline level seen with the pre-nominal structure. This may be due to the unnaturalness of the filler expressions, particularly the single-adjective phrases (e.g., "der Weinenden?" [the crying?]), which were initially designed to disrupt syntactic consistency in the Maze Task from the speaking side. To address this potential issue, Experiment 5 will replace these fillers with a more naturalistic production task.

5.2 Experiment 5

Experiments 1-4 have shown similar results: while a group of participants adopted a consistent syntactic structure for reference description, mainly the pre-nominal modification. Another group, particularly in the interactive setup of Experiment 4, varied the use of pre- and post-nominal modifications. These participants were more likely to vary their syntactic choices, tending to place the more informative property earlier in the description.

However, much remains unknown about how speakers would perform in face-to-face spoken communication. In the previous online experiments, participants were told they were collaborating with an online partner, but the partner was actually a simulated Confederate providing auditory descriptions and textual feedback. The plausibility of collaborating with a real-time online partner was limited due to factors like the use of synthesized speech and simulated response times. Additionally, participants' level of task involvement was harder to control when the task was performed outside of a controlled lab setting.

Moreover, the Maze Task used in these experiments presented all possible lexical options at each step, which helped to disentangle speakers' initial choices that determined property ordering. However, it remains unclear whether linearization choices observed in the Maze Task would align with those in spoken descriptions.

Another potential issue with the Maze Task is that it may have distracted participants' attention from the visual context and target figures. As the maze options were visually presented below the displays, participants would have to repeatedly shift their focus between the display and the maze, potentially disrupting their inspection of the displays. In contrast, a spoken description task would allow speakers to focus on the visual scene while producing descriptions, offering a more natural setup that maintains focus on the visual context.

Experiment 5 aimed to address these concerns by focusing on face-to-face communication, which is thought to provide the most engaging communication dynamics, compared with online communication (e.g., Kuhlen & Brennan, 2013; Peña et al., 2023). It introduced the trial-by-trial role alternation setup and the two critical conditions from Experiment 4 into a lab setting, where participants communicated face-to-face with a co-present Confederate: a well-trained native German speaker. Experiment 5 adopted a spoken production task, where participants and the Confederate alternated speaking and listening, allowing them to fully attend to the visual displays while speaking.

Experiment 5 aimed to investigate how participants would behave in this face-to-face spoken communication setting and whether their behavior would align with the patterns observed in the Maze Task in the previous experiments. The study explored whether a more natural and engaging setup in face-to-face communication could enhance speakers' sensitivity to informativity (i.e., RER) and affect their linearization choices.

If a similar pattern is observed in Experiment 5 as in Experiments 1-4, with one group of speakers varying their syntactic use and adopting the informative-first linearization preference, it would suggest that speakers' consideration of informativity in property ordering is not purely driven by the Maze Task. If these effects are further enhanced, such as detecting more frequent use of the informative-first linearization preference or a higher proportion of *Group Varied* Speakers compared to Experiment 4, it would suggest that the more engaging face-to-face environment encourages a listener-oriented approach in referential production. Alternatively, if participants' syntactic choices were distributed similarly across conditions, it would suggest that informativity does not play an essential role in determining property ordering in spoken production.

5.2.1 Method

5.2.1.1 Participants

45 native German speakers were recruited from Saarland University (29 females, mean age = 23.28). Each participant received 13 euros, based on the German minimum wage.

5.2.1.2 Conditions

The conditions for both participants and the Confederate, as well as their combinations, were designed in a manner similar to Experiment 4 (see Section 5.1.1.3). To briefly summarize, the two critical conditions in the Participant Turn, i.e., the Animal-Informative and Action-Informative Conditions, were equally combined with the two critical conditions (with their respective informative-first ordering structures) and filler conditions in the Confederate Turn. Table 5.4 summarizes the detailed combinations of conditions for the participants and the Confederate.

The only change in Experiment 5 was the replacement of the previous filler conditions with a new set of **Counting fillers**. These fillers required the interlocutors to communicate about the number of figures containing a shared property

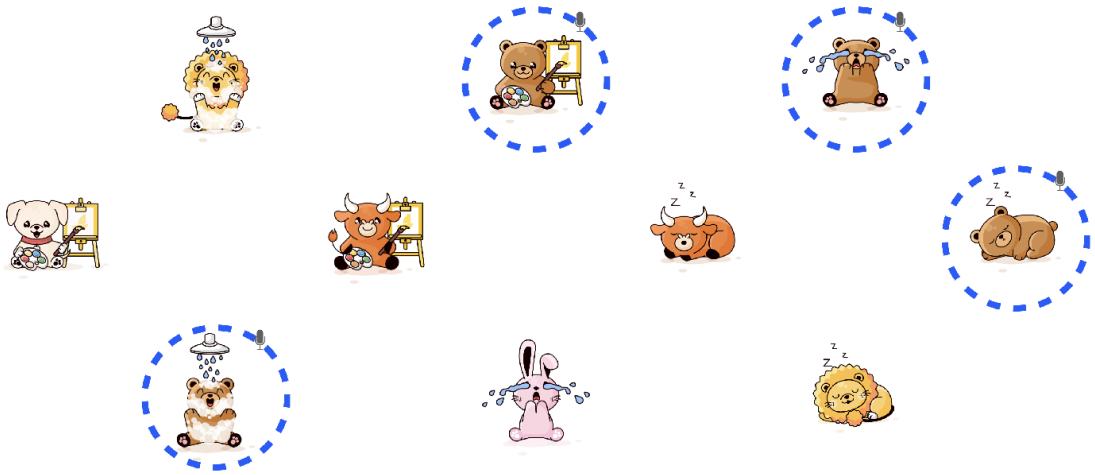


Figure 5.4: Example display for a counting filler condition.

in a display (see Figure 5.4). For example, the Participant would ask: "How many animals are bears?" and the Confederate would respond: "Four". There were two types of Counting fillers: **the Animal-counting Filler**, focusing on the number of figures with a certain animal property (e.g., "Wie viele Tiere sind Bären?" [How many animals are bears?]) and **the Action-counting Filler**, focusing on the number of figures with a certain action property (e.g., "Wie viele Tiere weinen?" [How many animals cry?]). The answers to these questions ranged from two to seven. The two filler conditions, along with their answers, were evenly distributed across a list.

The Counting fillers replaced the previous fillers used in Experiment 4 because the Maze Task fillers were not applicable in a spoken production context. The previous fillers involved a single adjective (e.g., "der Weinende?" [the crying?]), which is unnatural and rare in spoken communication. Additionally, the results from Experiment 4 indicated that a single noun has a similar priming effect as the pre-nominal modification, which may not have been sufficient to reduce the tendency for participants to stick to a consistent strategy. Apart from reducing the tendency for self-priming, the Counting fillers also aimed to direct participants' attention to the visual-numeric dimension of the display, which was relevant to the concept of RER and selection scopes.

5.2.1.3 Trial Orders and Lists

The trial orders, which primarily involved combinations of the Participant and Confederate Turns, were arranged in the same manner as in Experiment 4 (see Section 5.1.1.4). Experiment 5 included three types of experimental rounds (see Table 5.4):

1. Filler Round: The Confederate was presented with a critical condition, while the Participant encountered a filler condition.

2. Mixed Round: The Confederate was presented with a filler condition, followed by the Participant Turn with a critical condition.

Table 5.4: Experiment 5. Overview of condition combinations for the Confederate and Participant Turns across Round Types

Round type	Conditions: Confederate Turns	Conditions: Participant Turns
Critical Round	Animal-Informative (Post-nominal)	Animal-Informative
	Action-Informative (Pre-nominal)	Animal-Informative
	Animal-Informative (Post-nominal)	Action-Informative
	Action-Informative (Pre-nominal)	Action-Informative
Mixed Round	Animal-counting filler	Animal-Informative
	Action-counting filler	Animal-Informative
	Animal-counting filler	Action-Informative
	Action-counting filler	Action-Informative
Filler Round	Animal-Informative (Post-nominal)	Animal-counting filler
	Action-Informative (Pre-nominal)	Animal-counting filler
	Animal-Informative (Post-nominal)	Action-counting filler
	Action-Informative (Pre-nominal)	Action-counting filler

3. Critical Round: Both the Confederate and the Participant were presented with a critical condition.

This design ensured that the critical conditions in the Participant Turns were preceded by equal number of critical conditions in the Confederate Turn (and thus the same number of both syntactic structures) as well as by an equal number of filler conditions. At the same time, the combination of rounds ensured that the communication labor of being both a speaker and a listener was equally distributed between the Participant and the Confederate. The three types of Rounds were grouped into sets of six trials. In each group, the Filler Round was always presented first, and the order of the remaining two rounds was randomized. The groups, consisting of the three Rounds, were further randomized into a full experimental list.

In Experiment 5, there were two experimental lists. Each list contained 72 trials for the Participant Turn and 72 trials for the Confederate Turn. Within each list, 48 trials were the critical conditions in the Participant Turn. These 48 visual displays were expanded from the 24 displays used in the previous experiments, with the same counterbalancing criteria (see Chapter 4.1). Across an experimental list, each target (from the 24 unique figures) was presented twice, once in each critical condition. The stimuli used in the Confederate Turns and the Counting fillers in both turns remained the same across the two lists. The constraint was that there were no overlapping properties between the target properties in the Confederate Turns and those presented in the subsequent Participant Turn to avoid lexical repetition.

Table 5.5: Experiment 5. Overview of task assignments for the participants and the Confederate

	Confederate	Participant
Speaker Task	Click on the target (pre-defined)	Describe and verify
Listener Task	Describe and verify (pre-defined)	Click on the target

5.2.1.4 Procedure

The experiment was conducted in two connected lab rooms separated by a glass window. The participants were guided by the experimenter to meet and greet the female native German Confederate, and both were told that they were naive participants in the communication game. After the introduction, the participants and the Confederate were assigned to separate lab rooms.

Both the participants and the Confederate wore headphones and sat in front of a monitor paired with a microphone in their respective rooms. They could hear each other through the headphones and communicate via the microphones. The participants used a mouse next to the monitor, while the Confederate used a keyboard. Both devices were connected to the same computer. The experiment was implemented in LabVanced, the same as in the previous experiments.

Before starting the official experiment, both the participants and the Confederate completed three practice sessions. Session 1 and Session 2 introduced the counting filler trials and critical trials separately, while Session 3 mixed both types of trials on a trial-by-trial basis, as in the official experiment. The procedure in the practice trials was identical to the official experiment, except that if participants provided an incorrect answer, the experimenter gave reminders and corrections.

The entire experiment session lasted about an hour, with two breaks after every 24 experimental Rounds.

Task Procedure for Critical Conditions The joint goal of the critical trials was to determine whether the location arrangement of the ten figures was identical in the two displays viewed by both interlocutors. This goal was to be achieved by checking the location of the target figure. From the perspective of the naive participants, the general communication process for both parties should proceed as follows:

First, both parties viewed a display with ten animal figures for 4000 ms. Then, a red square appeared to highlight the target for the speaker, who would then ask the listener (who did not know the target) to click on the target figure, e.g., "Klicke auf den Hund, der duscht" [Click on the dog that showers]. The listener would then click on the corresponding target on the screen, and the chosen location would be marked with a red circle, visible to both interlocutors. Based on the feedback of the target location, whether it matched or differed from the target location in the speaker's display, the speaker would then verify if the location arrangement of the figures was identical or not.

In reality, however, the Confederate viewed the same screen as the participant but was trained to utter a list of pre-defined content based on the experimental conditions and to respond via the keyboard. Thus, the procedures for the Confederate differed from those of the participants.

To clarify how the seemingly symmetric communication process was established for the participants, two tasks are defined from the participant's perspective: **the Listener Task** and **the Speaker Task** (see Table 5.5 for an overview). The Speaker Task occurred when the participants were the ones asking the Confederate to click on a target (i.e., Participant Turn), while the Listener Task occurred when the participants were the ones clicking on the target based on the Confederate's description (i.e., Confederate Turn). The participant and the Confederate alternated roles in describing the target, so from the participant's perspective, the Speaker Task and Listener Task alternated trial by trial.

The Speaker Task (Participant Turn) In the Speaker Task, the participants initiated the conversation. In each trial, both parties first viewed a display with no indication of the target. After 4000 ms, a red square with a microphone icon highlighted the target figure for the participant, and a sound cue (a short ping sound) was played. Both the sound cue and the red square served as reminders for the speaker to begin speaking and ask the Confederate to click on the target figure. Their utterances were recorded.

The listener then clicked on the target, marked by a red circle with a mouse icon. The selected location was simultaneously displayed to the speaker. In reality, the location of the red circle feedback was pre-defined by the experimenter, and the Confederate only needed to press the Spacebar on the keyboard for the circle to appear.

The participant then verified whether the displays were identical by clicking on one of two buttons ("same" or "different") displayed below the figure. These buttons appeared 1000 ms after the location feedback. If the circle overlapped with the square, the participant should choose "same"; otherwise, they should select "different". In total, 1/3 of the speaker trials in the critical condition were verified as "different", where the red circle appeared in a different location from the target figure.

If the participant described the target incorrectly (e.g., "a showering rabbit" instead of "a dog"), the Confederate pressed the X button. This action always caused a figure that was not the target to be highlighted, and it appeared in a different location from the actual target. To increase authenticity, as suggested by Kuhlen & Brennan (2013), the Confederate was trained not to immediately press the keyboard button after the participant's description. Instead, they waited for one to two seconds to simulate the comprehension process.

The Listener Task (Confederate Turn) In the Listener Task, the participants responded based on the Confederate's description. In each trial, the speaker viewed the display without any indication of the target location. After 4000 ms, a sound cue (different from the Speaker Task) was played. The Confederate then began asking the participants to click on the target figure. In reality, the Confederate's descriptions were pre-defined, starting with the most informative property, leading to the use of pre-nominal modifications in the Action-Informative Condition and post-nominal modifications in the Animal-Informative Condition. The Confederate was encouraged to use pauses and hesitations (e.g., "uh," "um") before and during speaking to simulate the behavior of a naive participant.

The participants then clicked on the target figure on the screen, after which the red circle with the mouse icon appeared. The Confederate could then see the location of the figure the participants clicked on and provided feedback on whether the displays were identical or not. In reality, the verification feedback ("same" or "different"), presented below the display, was pre-programmed by the experimenter. One-third of the critical listener trials were programmed to show a "different" case. If the participants clicked on a mismatched figure, the verification was always "different".

Task Procedure for Filler Conditions The general procedure for a filler trial was similar to that of the critical trial, with the main difference being that more than one figure (ranging from 2 to 7) would be highlighted in a blue dashed circle after the preview. The speaker would then ask about the number of animals with a shared property, e.g., "Wie viele Tiere malen?" [How many animals paint?] . The listener's answer was provided orally. In the Speaker Task, after the participant's question, the Confederate's answer was pre-defined. One-third of the answers were not identical to the number of figures highlighted on the display, which should lead to a mismatch verification from the participants.

5.2.1.5 Analysis

The accuracy of each participant in the Listener Task and Speaker Task was first calculated. In the Speaker Task, a trial was considered accurate when the participant correctly described the target properties and, at the same time, correctly verified whether the display was identical to their partner's. In the Listener Task, a trial was accurate when the participant correctly responded by clicking on the matched target figure in the critical trials and by counting the number of animals correctly in the filler trials. The overall accuracy for the participants was reasonably high (above 87.5%), and as a result, no participants were excluded at this stage.

An additional 17 individual trials were excluded due to inaccurate descriptions (e.g., the participant used "dog" when "bear" was the target), technical issues, environmental interruptions, and syntactic repairs (when the participant clearly changed syntactic structures in the middle, such as switching from pre-nominal to post-nominal modification, e.g., "Klicke auf den mal... uh auf den Hase, der malt." [Click on the pain... uh on the rabbit that paints]).

Participants' spoken descriptions were transcribed into text and coded as pre-nominal or post-nominal modifications. Proportional data and their associated confidence intervals (CIs) were calculated using the same method as in Pre-test 1 (see Section 3.1.1.5).

Participants were categorized into *Group Varied* or *Group Consistent* based on the same binary criterion established in Experiment 1 (see Section 4.1.2.2). To test whether the proportion of *Group Varied* participants in Experiment 5 differed significantly from Experiment 4, a Two Proportion Z-Test was conducted, following the same method as in Experiment 2 (see Chapter 4.2).

Two logistic mixed-effects model analyses were conducted in the same sequence as in Experiment 4 (see Section 5.1). The first model included `Condition` as the primary predictor, following the model selection procedure used in Experiment 1 (see Chapter 4.1). The random effects structure of the best model included by-subject and by-item random intercepts.

The second model included the main effects of `Condition` (Animal-Informative and Action-Informative Conditions) and `Priming` as the control variable, where the primed structures in the Confederate Turn were either pre-nominal or post-nominal structures in the critical trials, and the filler questions (e.g., "How many animals are ___?") in the filler trials. Although the fillers may not elicit a specific syntactic priming effect (since they were not nominal phrases), they were still included in the analysis to maintain consistency of statistical models with Experiment 4. The contrast coding scheme for the two predictors was identical to that in Experiment 4. Random intercepts for subjects and items were included in the best model.

The timing of utterance planning was analyzed using First Property Latency (FPL), defined as the interval between the marking of the target and the onset of the first property. Full details and results of this exploratory analysis are provided in Appendix A.2.

5.2.2 Results

5.2.2.1 Group Proportions

Participants' consistency in the use of syntactic structures was first analyzed. Five participants (11.1%) consistently adopted a single structure throughout the experiment: four exclusively used the pre-nominal structure, while one exclusively used the post-nominal structure. The proportion of *Group Varied* participants in the current experiment (88.89%) was significantly higher than in Experiment 4 (64.56%), based on the Two Proportion Z-Test, $\chi^2(1, N = 124) = 8.69, p < .01$. Participants in *Group Consistent* (N = 5) were excluded from further analyses.

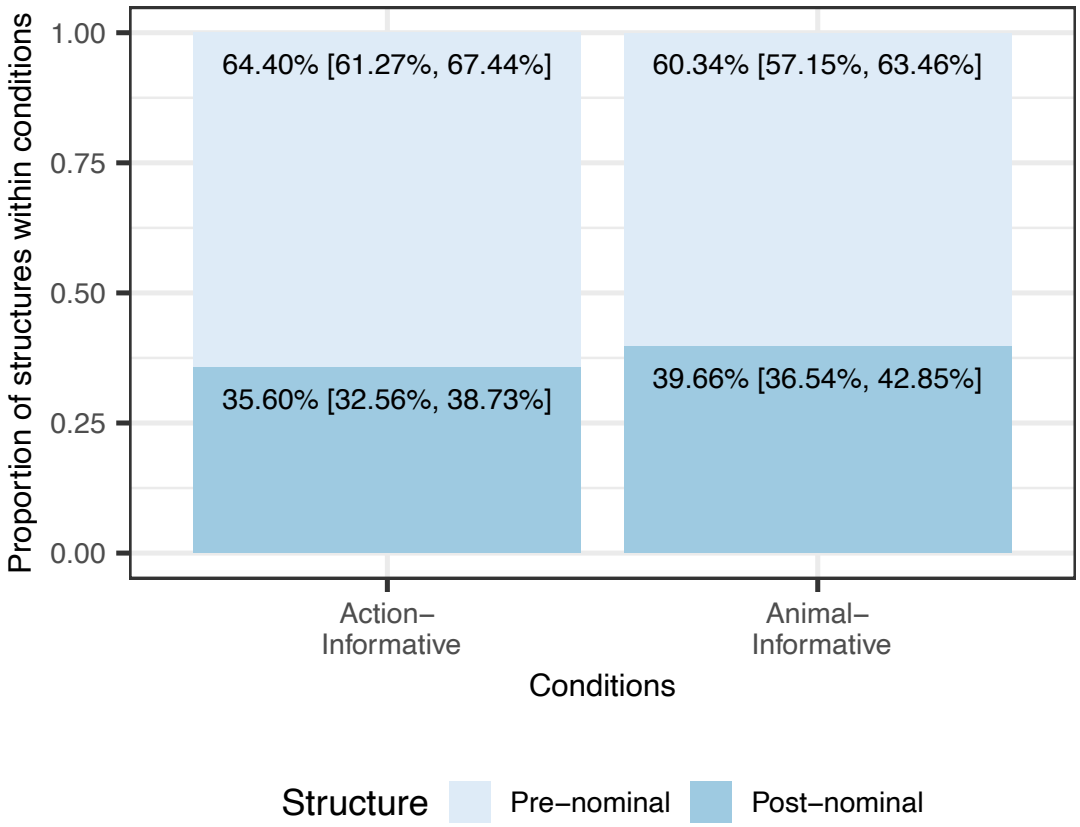


Figure 5.5: Experiment 5. the proportion of the use of pre-nominal (action-first) and post-nominal (animal-first) modifications within the two critical conditions. 95% CIs shown in brackets.

5.2.2.2 Results of *Group Varied*

Figure 5.5 presented the overall use of pre-/post- nominal structures in the two conditions. Although pre-nominal modifications were still preferred, post-nominal modifications were also more frequently used in general in the current experiment than in the previous experiments (62.37% pre-nominal vs. 37.63% post-nominal modifications, collapsed across conditions). The main effect of `Condition` was observed: the frequency of the use of post-nominal modifications was significantly higher in the Animal-Informative Condition (39.66%, proportion within condition), compared with the Action-Informative Condition (35.60%). ($\beta = 0.28, SE = 0.12, z = 2.40, p < .05$).

5.2.2.3 Results of Priming and Informativity

The results of the second model, which included `Condition` and `Priming` as predictors, are summarized in Table 5.6, and the effects are visualized in Figure 5.6. In addition to the main effect of `Condition`, a significant main effect of `Priming` was observed for every structure compared to pre-nominal modifications: participants were significantly more likely to produce post-nominal structures when

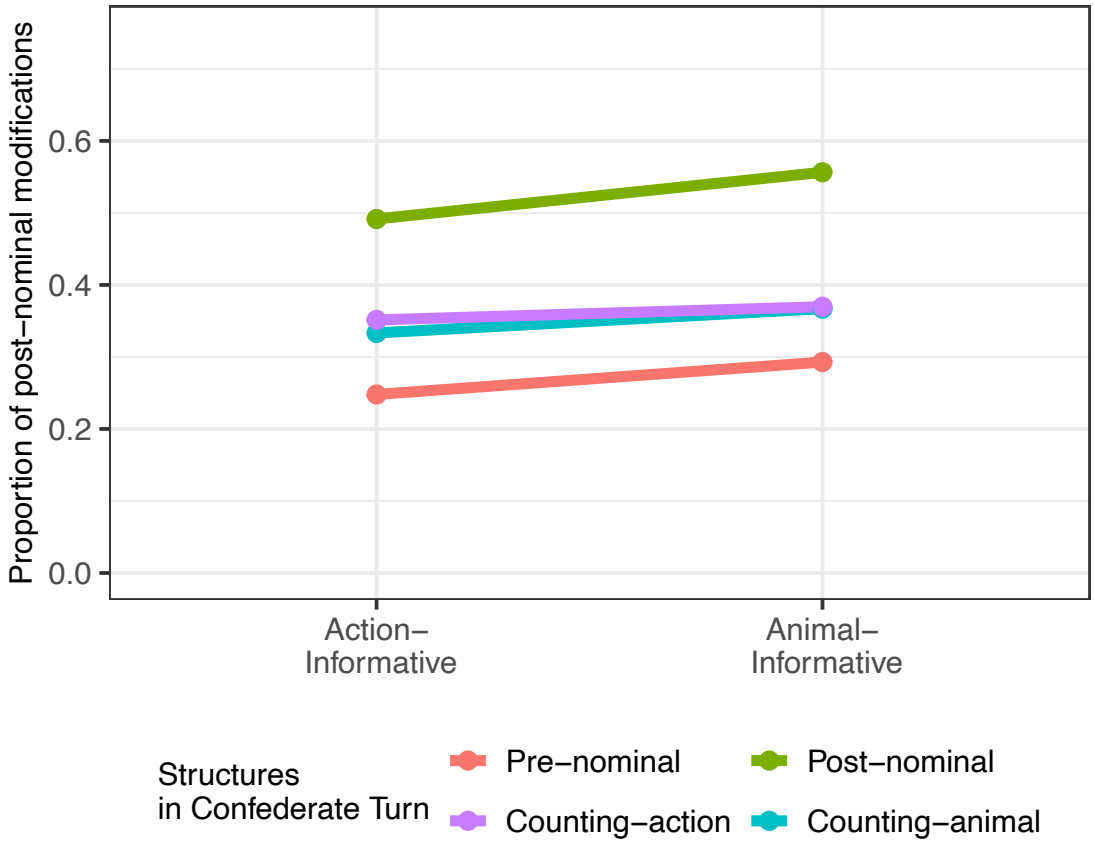


Figure 5.6: Experiment 5. Proportion of post-nominal modifications in each critical condition, grouped by the preceding syntactic structure used by the Confederate.

Table 5.6: Experiment 5. Results from the logistic mixed-effects model testing the main effects of Condition and Priming.

Predictor	$\hat{\beta}$	95% CI	z	p
Intercept	-1.45	[-2.01, -0.89]	-5.05	$< .001$
Condition: Animal vs. Action	-0.14	[-0.25, -0.02]	-2.34	.019
Priming: Post-nominal vs. Pre-nominal	1.66	[1.31, 2.01]	9.31	$< .001$
Priming: Counting-animal vs. Pre-nominal	0.58	[0.23, 0.93]	3.28	.001
Priming: Counting-action vs. Pre-nominal	0.64	[0.29, 0.98]	3.61	$< .001$

Note. The model was specified as: Structure ~ Condition + Priming + (1|subject) + (1|item). Condition (Animal-Informative and Action-Informative) was sum-to-zero coded. Priming (pre-nominal, post-nominal, Counting-animal, Counting-action) was dummy-coded with the pre-nominal structure as the reference level. Each priming coefficient represents a comparison to the pre-nominal baseline. "Counting-animal" and "Counting-action" refer to filler turns where participants were prompted to count the number of animals or actions, respectively.

they had previously been primed with a post-nominal structure in the Confederate Turn. The frequency of post-nominal structure use also increased significantly after participants encountered the filler conditions in the Confederate Turn, compared to when they had previously been primed with a pre-nominal structure.

A second model, which included the interaction between `Condition` and `Priming`, was also tested. However, this interaction model did not significantly improve the model fit compared to the model reported in Table 5.6 ($BIC_{main} = 1927.32$, $BIC_{interaction} = 1949.14$, $\chi^2(3) = 0.84, p = 0.84$).

5.2.3 Discussion

Experiment 5 examined the impact of informativity on speakers' linearization strategies in spoken production during face-to-face communication. Participants interacted with a Confederate in a lab setting, alternating target descriptions on a trial-by-trial basis. This setup aimed to investigate whether the interactive and engaging nature of face-to-face communication would further enhance the use of the informative-first linearization preference, compared to the online communication settings in previous experiments.

The main findings replicated the results of the Maze Task in previous online experiments. Speakers in *Group Varied* were more likely to place the more informative property first when comparing the two critical conditions, reflecting the tendency toward the informative-first linearization preference. Additionally, as observed in Experiment 4, both priming and informativity significantly influenced property ordering without statistical interaction. Experiment 5 further validated that the modulation effect of informativity was not purely induced by the Maze Task, but rather, informativity also affects property ordering in spoken referential production.

Interestingly, Experiment 5 recorded a notably higher proportion of post-nominal structures (37%) compared to previous experiments (approximately 20%). This distribution of pre- versus post-nominal usage closely resembled the distribution observed in Pre-test 1 (39%) introduced in Chapter 3.1, which also employed a spoken production task.

One potential explanation for the different proportional use of the two structures between the spoken production and Maze Tasks could be the varying levels of formality associated with spoken and written modalities in German. For the current set of stimuli, the pre-nominal adjectives, known as "Partizip I" [Participle I], are formed by an infinitive verb plus the suffix "-d" (e.g., the adjective "weinend" [crying] derived from the infinitive verb "weinen" [to cry] with the addition of "-d"). These participle constructions are often considered more formal in style and are therefore more common in written texts than in spoken speech (e.g., Mattmüller, 2016; van Miltenburg et al., 2018). Although the Maze Task used in this project was set in a communicative context, it may still have been perceived as a form of written expression, contributing to the preference for the more formal pre-nominal structures by native speakers.

More importantly, the proportion of *Group Varied* speakers increased even further in Experiment 5 (almost 90%) compared to Experiment 4 (approximately 65%), adopting the same trial-by-trial role alternation. This suggests that the

growth of *Group Varied* may be not solely influenced by syntactic priming, but also that the level of experimental engagement in face-to-face communication and the spoken production task contributed to speakers' sensitivity towards informativity and so their linearization strategies. Most participants in Experiment 5 were able to vary their use of syntactic structure, forming the tendency of the informative-first linearization preference. The General Discussion below will further compare these results with previous experiments.

Another observation concerns the priming effects. The results were similar to those of Experiment 4, with both priming and informativity significantly affecting speakers' syntactic choices, without a statistical interaction. Particularly in the Critical Rounds, speakers tended to repeat the structure used by the Confederate in the preceding trial (e.g., producing more post-nominal modifications after being primed by post-nominal structures). At the same time, they were also sensitive to informativity, using more post-nominal modifications in the Animal-Informative Condition compared to the Action-Informative Condition. The General Discussion below will further explore these results.

5.3 General Discussion

Experiments 4 and 5 examined how different communication dynamics influence the effect of informativity on speakers' linearization strategies. Communication dynamics were manipulated along two dimensions: 1) the level of perspective-taking, enhanced through trial-by-trial role alternation, and 2) the degree of communication engagement, comparing online communication in Experiment 4 with in-lab, face-to-face communication in Experiment 5. Both manipulations aimed to determine whether the informative-first linearization preference, as a listener-oriented strategy, is strengthened in the more interactive and engaging communication settings.

A secondary investigation focused on the interplay between the effects of priming and informativity, as trial-by-trial role alternation inherently introduced priming. The key question was whether, and to what extent, informativity would continue to influence property ordering when disentangled from the effects of priming.

Both experiments replicated the main findings from Experiments 1-3 presented in the previous chapter. A subset of speakers, identified as *Group Varied*, showed the tendency of prioritizing the more informative property word in referential expressions when comparing the two critical conditions, resulting in an informative-first linearization strategy. This pattern supports the listener-oriented view of referential production, as presenting the more informative property word with a higher RER earlier in the expression narrows down potential target candidates more efficiently, facilitating the listener's target search.

Experiments 4 and 5 further revealed that the proportional distribution of *Group Varied* and *Group Consistent* speakers is not static. Instead, the proportion of *Group Varied* speakers increased significantly in both Experiments 4 and 5 compared to Experiments 1-3, where participants performed only as speakers within blocks. Notably, the size of *Group Varied* was further enhanced in the more engaging and authentic face-to-face communication context of Experiment 5

compared to online communication in Experiment 4. In contrast, the proportion of *Group Consistent* speakers decreased in these two experiments compared to Experiments 1-3, where these speakers adopted a speaker-oriented, heuristic strategy by adhering to the same syntactic structure throughout an experiment.

Both experiments also identified independent main effects of informativity and priming. These findings suggest that property ordering is influenced by both the tendency of structural alignment between interlocutors and, more importantly, the informativity profile quantified by RER.

The general discussion below will focus on the effects of the detailed experimental manipulations introduced in Experiments 4 and 5, namely, trial-by-trial role alternation, online versus face-to-face communication contexts, and priming. A more theoretical discussion of these results will be presented in Chapter 6.

5.3.1 Trial-by-trial Role Alternation

The growth of *Group Varied* in Experiments 4 and 5, compared to Experiments 1-3, supports the idea that more engaging and interactive communication environments encouraged a greater number of speakers to employ property ordering sensitive to informativity, reflecting a listener-oriented approach in referential production. Communication engagement may be enhanced by the trial-by-trial role alternation design employed in these experiments, because it reinforced frequent perspective changes by requiring participants to switch between speaker and listener roles on a trial-by-trial basis. This setup allowed participants to experience the incremental comprehension process of target search firsthand, making them more likely to consider the listener's need of efficient target search when planning their referential expressions (see Chapter 2.1.2.1 and Chapter 2.2.2.1 for a detailed discussion of incremental comprehension).

As discussed in Chapter 2.2.2.4, this design represents a stronger, explicit manipulation for triggering perspective-taking by compelling speakers to adopt the perspective of their addressees. This manipulation elicited a more robust listener-oriented effect compared to other relatively moderate perspective-taking manipulations, such as presenting a separate Listener Block in Experiment 2 (see a discussion in Chapter 4.4.3).

Employing different experimental manipulations to induce moderate versus strong demands on perspective-taking is also common in research using the director-matcher tasks, where speakers are exposed to privileged knowledge inaccessible to listeners. In these tasks, the key measure of interest is the extent to which speakers include or exclude privileged knowledge in their utterances when such knowledge is irrelevant to the listeners. This reflects the degree to which speakers adopt an egocentric perspective in their referential descriptions.

Yoon et al. (2012) investigated perspective-taking demands by comparing different communication tasks and goals for speakers. Specifically, they contrasted the task of *informing*, describing changes in a display made by a third party to the listener (e.g., "the experimenter will move the big plate to the left"), with *requesting*, which involved directly guiding the listener to move objects in the display (e.g., "could you move the plate to the left?"). *Requesting* was considered a

stronger demand for perspective-taking, as it required speakers to take the listener's perspective to guide the action, whereas *informing* merely described a situation. Yoon et al. (2012) found that speakers more frequently omitted the privileged knowledge during the *requesting* task than the *informing* task, suggesting that when communication tasks and goals emphasize the perspective of listeners, speakers are more capable of perspective changing to tailor their utterances for listeners by including properties only relevant to listeners' tasks.

However, not all explicit manipulations for perspective-taking in previous research successfully elicited listener-oriented production outputs. For instance, Damen et al. (2019) attempted to stimulate speakers' attention toward listeners' perspectives by explicitly framing questions from either an "other-focus" perspective (e.g., "Which figures are visible to your addressee?") or a "self-focus" perspective (e.g., "Which figures are visible to you?"). In their second experiment, speakers were even required to physically move the corresponding figures to answer these questions. Despite these explicit manipulations before the target description in each trial, speakers' inclusion of privileged knowledge in referential expressions remained unaffected across conditions, similar to their baseline level without these questions concerning perspectives. This suggests that even when speakers are made explicitly aware of the addressee's perspective in contrast with their own perspective, such awareness does not necessarily influence property selection decisions encoded in their utterances.

These detailed experimental manipulations for triggering perspective-taking, e.g., trial-by-trial role alternation in Experiments 4 and 5, introducing different communication tasks and goals for speakers (Yoon et al., 2012), and asking explicit questions to direct speakers' attention of self- or other-focused perspectives (Damen et al., 2019), present a complicated picture of perspective-taking and its effects on property selection and property ordering.

First, perspective-taking is not an automatic or static process in communication. The extent to which speakers engage in perspective changing towards addressees would depend on specific experimental manipulations and task demands.

Second, the relationship between perspective-taking and speakers' production outputs is not always related. While it is certain that speakers are explicitly aware of and capable of perspective-changing in referential communication tasks, also evidenced by speakers' neural activities detected via fMRI (Vanlangendonck et al., 2018), this does not always lead to changes to encoding strategies in referential expressions to meet addressees' needs. While supporting evidence has been found in the current Experiments 4 and 5, as well as findings by Yoon et al. (2012), suggesting that speakers actively adjust their production strategies, Damen et al. (2019), in contrast, indicates that such adjustments are not guaranteed, even with strong demands on perspective-taking in the task, let alone with other moderate manipulations such as the block design in Experiment 2 and in Vogels et al. (2020) (see Chapter 2.1.3.1). A more theoretical discussion regarding the relationship between perspective-taking and listener-oriented encoding strategies will be revisited in Chapter 6.2.2.

5.3.2 The Function of Confederate in Online vs. Face-to-Face Communication

The different proportions of *Group Varied* in Experiments 4 and 5 highlighted the impact of communication engagement possibly influenced by perceived authenticity of the Confederate and/or by different types of communication dynamics, i.e., online versus face-to-face communication. These two aspects are discussed below.

One interpretation of the effect of communication engagement involves the perceived authenticity of the Confederate, specifically, the extent to which participants believed they were interacting with a genuine partner. In Experiment 4, this was assessed through two rating questions in the final questionnaire of the experiment: 1) the perceived cooperativeness of the partner ("How cooperative do you think your partner was during the experiment?") and 2) the extent to which participants thought the partner was human or virtual. Participants rated the partner as highly cooperative (mean score: 1.2 on a 4-point German scale, where 1.0 indicates the highest level of collaboration) but predominantly perceived the partner as a virtual agent (mean score: 3.6/4).

In contrast, feedback from Experiment 5 indicated higher perceived authenticity of the Confederate. Only one out of 45 participants suspected, but was not certain, that the partner was also an experimenter, while the majority believed they were interacting with another naive participant. This suggests that face-to-face interactions in Experiment 5 significantly enhanced the perceived authenticity of the Confederate, contributing to a more engaging communication environment that led to more listener-oriented encoding strategies in property ordering.

Kuhlen & Brennan (2013) categorized several types of dialogue partners and their different functions as Confederates in experiments involving dialogues, emphasizing the impact of the Confederate's behavior on experimental results. Based on their categorizations, the partner in Experiments 4 and 5 can be identified as both "the collaborative partner" and, to some extent, "the interactively aligned partner".

The Confederate functioned as "a collaborative partner" because this role in Experiments 4 and 5 went beyond being a passively co-present addressee. Instead, the Confederate actively contributed to achieving communication success in the referential communication task, both as a speaker in target descriptions and as a listener in target identification. However, the Confederate functioned only partially as "the interactively aligned partner". This is because, on the one hand, the experiments predicted the priming effect as a natural tendency for alignment in interaction (Pickering & Garrod, 2004). On the other hand, however, this alignment was not reciprocal between the two interlocutors. Instead, alignment occurred only for the participants, who were aligned with the Confederate, but not the other way around, because the Confederate also served as the "scripted Confederate", as described by Kuhlen & Brennan (2013).

The inclusion of a Confederate in the current project served several critical functions. Beyond facilitating collaborative interactions in both online and in-lab experiments, the Confederate's pre-defined utterances were designed to consistently adhere to the informative-first linearization preference. This design aimed to create an efficiency-driven communication environment (see Chapter 1 for details) while

also ensuring a balanced distribution of syntactic structures preceding the critical conditions, crucial for systematically measuring the priming effect under controlled conditions. Consequently, the Confederate's speech was "scripted" rather than spontaneous.

However, Kuhlen & Brennan (2013) raised concerns regarding the reduced naturalness of speech from "the scripted Confederate", as such utterances might be perceived as "pragmatically infelicitous" or inauthentic, potentially affecting participants' behavior. To mitigate these concerns, Experiment 5 trained the Confederate to incorporate elements to enhance the perceived naturalness of the Confederate's speech, including spontaneous speech errors and repairs, hesitation fillers like "uh," and natural pauses before speaking, which were also recommended acts for naturalness by Kuhlen & Brennan (2013). The synthesized speech in Experiment 4 was also designed with tonal features and pauses to increase naturalness and processing time for comprehension. Taking these together, while the Confederates in the current chapter were scripted and controlled, they were also perceived as an authentic, collaborative conversation partner that successfully increased communication engagement by increasing the proportion of speakers belonging to *Group Varied,* compared with experiments in the previous chapter.

A second aspect for interpreting the impact of communication engagement involves comparing the two communication dynamics between online and face-to-face communication. Peña et al. (2023) compared the influence of different communication dynamics using the director-matcher task in web-based experiments. They demonstrated that the framing of the "computer partner" significantly influenced speakers' egocentricity during HCC. In their first experiment, the computer partner was conceptualized as a simple computer machine producing synthetic speech, leading to higher egocentric production, as evidenced by more frequent inclusion of privileged knowledge in referential expressions compared to interactions with a human partner. In the second experiment, the computer partner was explicitly designed as a "virtual voice agent" with a simulated identity, resulting in encoding patterns similar to those observed in HHC. This suggests that perceiving the computer as a distinct dialogue partner motivates speakers to be more engaged in perspective-taking and to tailor their utterances for the partner's needs.

The simulated Confederate in Experiment 4, introduced in the current chapter as well as in Experiments 2-3, aligns with the framing of the second experiment in Peña et al. (2023). Although the experimental instructions introduced the partner as another human participant connected online, participants were prone to perceive it as a separate virtual voice agent (as suggested by the rating of the questionnaire), possibly due to artificial features such as relatively stable reaction times and consistent tonal variations in synthesized speech. This conceptualization of a separated voice agent would also contribute to the increased proportion of *Group Varied* compared to interactions with the more computer-like partners in Experiment 1 and the Pre-tests, where responses were purely textual.

However, unlike the indifferent spoken patterns between HCC versus HHC reported in the second experiment of Peña et al. (2023), Experiment 5 observed an even higher proportion of *Group Varied* in face-to-face communication compared to the online setup in Experiment 4. This discrepancy is likely due to differences

in the nature of experimental design: the HHC condition in Peña et al. (2023) adopted human pre-recorded audio descriptions, but the experiments were still conducted in an online format without physical co-presence of the partner (Kuhlen & Brennan, 2013). Experiment 5, in contrast, involved in-person, face-to-face interaction, which is supposed to be even more engaging than interacting with another human interlocutor via computers.

These observations suggest a gradient spectrum of communication engagement influencing perspective-taking and allocentricity (i.e., being listener-oriented). This spectrum ranges from the more egocentric interaction with a simple computer, to an intermediate level of allocentricity when the computer is perceived as a distinct dialogue partner, and finally to the most listener-oriented and engaging context of face-to-face communication.

With the rapid expansion of diverse communication interfaces (e.g., Cowan et al., 2023), HCC has emerged as a distinct field of study. While the current project was not initially designed to address this field but was primarily motivated by the need to create naturalistic and engaging online communication experiments during the COVID-19 pandemic, the results and experimental designs present relevance to this field. A broader discussion and review of HCC will be revisited in Chapter 6.3.2.3.

5.3.3 Priming

The last point for discussion is the replicated priming effect, which did not interact with the main effect of informativity, detected in both experiments. Pickering & Garrod (2004) proposes "the interactive alignment account", arguing that interlocutors in conversations tend to be automatically aligned across multiple levels of linguistic representations. The priming effect is one example of alignment. Pickering & Garrod (2004) argues that the tendency of alignment created a simplified processing routine for language production. Utterances produced by one of the interlocutors would already activate the syntactic representation for the other interlocutor, increasing the chance of revisiting the same production route and reusing the same syntactic structure. For the case of the Experiment 5 especially, this may account for the increased overall use of post-nominal modifications, possibly by reducing the increased production difficulty through this alignment process, where the more difficult post-nominal structure was previously activated for the participants in the Confederate Turn (see also Chapter 2.2.2.5).

Consistent with previous experimental results such as Fukumura (2018) and Haywood et al. (2003), who also studied informativity and priming jointly, there was no interaction between the two factors in deciding property ordering. A possible explanation for the orthogonal relationship of the two factors is that priming can be considered as a speaker-internal factor but informativity as a speaker-external factor (see Chapter 1.3 for definitions of these terms). Priming is speaker-internal because, as briefly discussed by Pickering & Garrod (2004), the alignment effect in dialogue may modify the default production routine as described in Levelt (1989)'s language production model (see Figure 1.3 in Chapter 1.3.1), i.e., from message encoding, to grammatical encoding, and to phonological encoding. Rather, an

utterance can be "built around" a previously activated linguistic representation, activated and primed by the other interlocutor. In the current case, the activated linguistic representation is the syntactic structure for target description. While the more detailed production mechanism and time course of production planning is beyond the scope of the current project, this interpretation, overall, considers priming a speaker-internal factor.

On the other hand, informativity is a speaker-external, communication-based factor because it is motivated by a listener-oriented goal of facilitating the listener's target search (see Chapter 1.3.2.2). This factor may affect property ordering in parallel with the speaker-internal factors such as priming, operating jointly during the grammatical encoding stage of language production. A more theoretical discussion of the effects of speaker-internal and speaker-external factors will be discussed in Chapter 6.2.2.2.

5.3.4 Conclusion

In sum, Experiments 4 and 5 demonstrated that a more interactive and engaging communication environment can encourage more speakers to be listener-oriented and adopt the informative-first linearization preference. Compared to Experiments 1-3, a larger proportion of speakers in *Group Varied* adjusted their syntactic choices based on informativity, aligning with this preference. The next chapter, Chapter 6, situates these findings within the broader context of communication efficiency and reference production.

6

General Discussion and Conclusion: Informative Linearization in Efficient Reference Production

Contents

This final chapter begins by summarizing the project presented in the preceding chapters. It then situates the findings within the broader research context of communication efficiency and language production from a theoretical perspective. Three main points are proposed: first, the informative-first linearization preference in the current work provides solid evidence in support of communication efficiency; second, perspective-taking is a mediating factor that influences the extent to which speakers adopt either a speaker-oriented or listener-oriented strategy in reference production; and third, the concept of communication efficiency is revisited, calling for a more precise definition of this term in research of language production and referential communication. The second half of the chapter addresses the limitations of

the referential communication task and suggests three directions for future research.

6.1 Project Summary

This thesis project set out by conceptualizing the referential communication task as an instance of a cooperative and efficient information transmission system (Chapter 1.1). Within this framework, the project investigated whether speakers' referential production is influenced by informativity in the pursuit of communication efficiency. Informativity is considered a speaker-external, communication-based factor, as being informative inherently reflects speakers' listener-oriented intention to facilitate listeners' identification of the target. Specifically, the project examined whether efficiency, and so the modulation effect of informativity, can be reflected in the linearization of referential properties that leads to syntactic variations in referential encodings.

Chapter 2 reviews previous research on property selection and property ordering in reference production, highlighting the motivation for studying informativity and property ordering as a window to disentangle the speaker-oriented and listener-oriented views, shifting away from a predominant focus on property selection, particularly overspecification (OS) in previous research. While research on OS has yielded mixed results and interpretations in distinguishing the two diverging views, examining property ordering and informativity would potentially provide a clearer picture.

Previous research has shown that speaker-external factors are generally less influential in determining syntactic choices during language production compared with speaker-internal factors such as grammatical and semantic conventions or priming effects (see Chapter 1.3.2.1 and Chapter 2.2). If informativity can nonetheless modulate property ordering at the syntactic level, it would strongly support the listener-oriented view that speakers seek to play a cooperative, efficiency-motivated role in referential communication by tailoring their utterances, more specifically, the order of referential properties, to facilitate listeners' target identification.

The current project investigates the effect of informativity on property ordering by identifying three related research questions. First, it examines whether informativity influences property ordering at the syntactic level at all. If any effect can be observed, the next question addresses the specific linearization pattern preferred by speakers. Finally, if a consistent linearization pattern modulated by informativity can be identified, the third question explores whether more interactive and engaging communication dynamics would further amplify the effect strength of informativity.

These research questions were addressed through five experiments presented across two chapters. Informativity was quantified using Referential Entropy Reduction (RER), a measure of the degree to which a property word reduces uncertainty in identifying the target figure in the visual display. The experiments examined the syntactic structures of pre-nominal and post-nominal modifications in describing stimuli featuring animals performing various actions. As shown in the Pre-tests (Chapter 3), these structures can be flexibly encoded in German, though with an overall preference for pre-nominal modifications. Each experiment compared two

critical conditions: the Animal-Informative Condition and the Action-Informative Condition. In both conditions, the property word with a higher RER is more informative, narrowing the selection scope to two figures within ten-figure visual displays. This was contrasted with the less informative property, which reduced the selection scope to five figures, creating larger uncertainty for referring to the targets.

Experiments 1–3, outlined in Chapter 4, investigated the first two research questions: whether and how informativity influences property ordering at the syntactic level. These experiments employed online communication games featuring a maze-based sentence completion task, designed to highlight the RER dimension of the visual displays, and to capture speakers' initial choice of the property that further determines the order and syntactic structure of the entire referential expression. To ensure participants' attention to the task, Experiment 2 introduced a Listener Block before the Speaker Maze Task, and Experiment 3 further animated the visual representations of the ten figures, to encourage active involvement and careful previewing of displays before reference production.

Across all three experiments, a consistent pattern emerged: one group of speakers, *Group Varied*, demonstrated sensitivity to informativity when comparing the two critical conditions, by preferring the structural encoding that places the more informative property word earlier in referential expressions, adhering to the informative-first linearization preference. This informativity-driven linearization preference influences property ordering at the syntactic level, particularly evidenced by the more frequent use of the less preferred post-nominal modifications (i.e., properties ordered as animal-action) in the Animal-Informative Condition than in the Action-Informative Condition. In contrast, pre-nominal modifications (action-animal), which was preferred in general, were more frequent in the Action-Informative Condition than in the Animal-informative Condition. The other group of speakers, *Group Consistent*, consistently adhered to a single syntactic structure, mainly pre-nominal modifications, throughout the experiment, demonstrating insensitivity to the manipulation of RER. The results of Experiments 2 and 3 replicated, but did not improve, the results of Experiment 1, indicating that the additional manipulations of introducing a Listener Block and display animation did not significantly enhance participants' sensitivity to RER.

Experiments 4 and 5, presented in Chapter 5, examined the third research question: whether communication engagement enhances the use of the informative-first linearization preference. These experiments employed trial-by-trial role alternations to reinforce perspective-changing and compared online communication paired with the maze task (Experiment 4) with face-to-face communication involving spoken production (Experiment 5).

While both experiments replicated the main effect of informativity observed in Experiments 1-3, they also revealed a significant increase in the proportion of *Group Varied* speakers, particularly in the face-to-face communication setting of Experiment 5. This suggests that higher communication engagement may encourage more speakers to adopt the informative-first linearization strategy. At the same time, both experiments showed the syntactic priming effects: speakers tended to align with the structure used by the Confederate in the preceding trial. Importantly, statistical analyses within the *Group Varied* speakers in Experiments

4 and 5 indicated that both informativity and priming exerted independent main effects on property ordering without statistical interaction. Thus, while priming also contributed to the increased proportion of *Group Varied* speakers, the results support the hypothesis that communication engagement plays a role in increasing speakers' overall sensitivity to informativity when planning referential expressions.

In conclusion, Experiments 1- 5 provided a clearer understanding of speakers' property ordering strategies in referential communication tasks. While one group of speakers tends to adopt a speaker-oriented approach, remaining insensitive to informativity and adhering habitually to one syntactic structure for encoding properties, speakers belonging to *Group Varied* support the listener-oriented view, seeking communication efficiency by adopting the informative-first linearization preference to facilitate listeners' target search. Notably, the proportion of listener-oriented speakers (i.e., *Group Varied*) increases when the communication tasks involve perspective-changing and a higher level of communication engagement. Below, these findings are further discussed in the context of communication efficiency and language production.

6.2 Project Discussion

6.2.1 Informative-First Linearization Preference as Evidence for Communication Efficiency

The current project consistently observed the informative-first linearization preference across five experiments involving both the maze-based sentence completion task (Experiments 1-4) and the spoken production task (Experiment 5) for *Group Varied* speakers, with the tendency of prioritizing the more informative property earlier in referential expressions when comparing across the two critical conditions, leading to syntactic differences given the current set of stimuli.

The informative-first linearization preference observed in the current project highlights the following three key aspects of how informativity supports communication efficiency, which are further discussed in this section: First, informativity is a speaker-external, communication-based factor that guides property ordering through "linear modification" (Fukumura & Santesteban, 2017). Second, the underlying motivation of seeking communication efficiency via informativity is a listener-oriented goal in referential communication tasks, where speakers from *Group Varied* assume listeners' incremental visual search processing, and exploit property ordering to facilitate this process for target identification. Third, informativity in this project, quantified by RER, is also in line with the notion of informativity in information-theoretic accounts, where the RER profile of the property words predicts the order of properties encoded in speakers' production outputs.

First, informativity guides property ordering by favoring the earlier placement of the more informative property in referential expressions. Similar linearization patterns modulated by informativity have been observed by Fukumura (2018) for adjective ordering, where the more discriminating property is mentioned first more often (e.g., "a *striped* green bowtie" when the "striped" pattern is more

discriminating), and by Haywood et al. (2003), where sorting priority (Color or Pattern box) modulates property ordering, even at the syntactic level in their Experiment 2 (e.g., "a *square* that's red" when sorting the Pattern Box, see Chapter 2.2.2.3 for more details of the two studies). In a supplementary study of Fukumura (2018), Fukumura & Santesteban (2017) adopted similar visual conditions as in Fukumura (2018) for Basque speakers, to further show that informativity affects modifier ordering in a language preferring post-nominal modifications. For both English and Basque, i.e., in both syntactic structures, the more discriminating property is preferred to be placed earlier.

These findings, together with the results of the current project, further indicate that informativity is qualitatively different from other conventional constraints and semantic-based factors in deciding property ordering (e.g., absoluteness, Martin, 1969; subjectivity, Scontras et al., 2017, other factors are introduced in Chapter 2.2.1). These semantic factors often align with the mirror-image ordering principle predicting a reversed ordering preference for post-nominal modifiers (e.g., Cinque, 1994).

In comparison, the effect of informativity is shaped by visually-situated contexts and is employed by speakers to facilitate efficient target search. Consequently, regardless of the conventional constraints on modifier ordering in pre- or post-nominal structures, informativity guides property ordering *linearly* (termed "linear modification" by Fukumura & Santesteban, 2017), rather than centering around the head noun and forming a mirror-image order in post-nominal positions ("hierarchical modification"). The current project extends the linear ordering effect of informativity beyond modifier ordering, where the position of the head noun was not focused, into syntactic ordering, where prioritizing the more informative property can lead to alternations between syntactic structures that determine the relative position of the head noun and its modifier. Notably, this pattern of linear modification should invite further replication in other languages that favor post-nominal modifications, such as Hebrew (e.g., Trainin & Shetreet, 2021).

Second, the informative-first linearization preference, observed in both modifier and syntactic ordering, highlights that informativity is an efficiency-driven, speaker-external factor, supporting the listener-oriented view (see Chapter 2.1.2), especially the Incremental Efficiency Hypothesis (Rubio-Fernández, 2016). The Incremental Efficiency Hypothesis, initially proposed to account for OS, suggests that speakers take into account an important aspect of common ground knowledge during the planning of referential expressions: listeners process visual search incrementally, relating each property word to the visual context in the order it is mentioned (see also Rubio-Fernández et al., 2021 and Chapter 2.2.2.1 for incremental comprehension). Evidence in previous research on property selection for this hypothesis includes, for instance, the less frequent use of post-nominal OS in Spanish compared to the more frequent pre-nominal OS in English (Rubio-Fernández, 2016).

The informative-first linearization preference provides further evidence for this hypothesis by demonstrating that property ordering is directly influenced by the informativity profile of RER in *Group Varied*. Speakers formulate property ordering incrementally by each property unit, identifying the more informative property word over the less informative one, and placing it earlier in referential expressions

to facilitate listeners' early target identification. This requires speakers in *Group Varied* to not only be sensitive to informativity and RER in the visual contexts, but also to consider the presence of listeners and the incremental comprehension processes shaped by the linear order of utterances.

Notably, this occurs even when adopting such ordering preferences may require additional production effort, as it involves inverting the preferred word order. Examples include producing post-nominal modifications in the Animal-Informative Condition in the current project and adopting the less preferred Pattern-Color adjective order when Pattern is more discriminating, as observed in Fukumura (2018). These evidence aligns with the listener-oriented view and audience design (Clark & Murphy, 1982), which will be further discussed in the next Section 6.2.2.

Third, the informative-first linearization preference is also consistent with predictions from several information-theoretic accounts, particularly Rational Speech Act (RSA) models (Cohn-Gordon et al., 2019; M. C. Frank & Goodman, 2012; Waldon & Degen, 2021, introduced in Chapter 2.2.3), which emphasize the role of informativity in property selection and ordering. The current project extends the RSA speaker model by moving beyond single-word utterances, showing that informativity also guides the ordering of multiple properties in referential production. While the RSA model predicts that the most informative word is the most likely to be chosen when limited to one-word utterances (M. C. Frank & Goodman, 2012), the current findings suggest that, without such constraints, informativity also affects the probability of which property is mentioned first, further shaping both the linearization and syntactic structure of referring expressions.

Importantly, informativity is not the only factor influencing property ordering. Word order preferences also exert a strong influence, as evidenced by the overall dominance of pre-nominal modifications across conditions in this project. Structural priming also plays a role in shaping property ordering, as speakers tend to reuse recently encountered syntactic patterns (see Chapter 2.2.2.2 and Chapter 5). Future computational models of property ordering may benefit from integrating these effects alongside informativity-based accounts.

Another comment concerns the lack of evidence relevant to the Uniform Information Density (UID) hypothesis (A. F. Frank & Jaeger, 2008; Jaeger, 2010, introduced in Chapter 1.4.1). UID proposes that speakers prefer utterances that distribute informativity uniformly across linguistic units to avoid processing peaks. Applied to property ordering, UID would predict a preference for placing the less informative property first, thereby smoothing the distribution of informativity across an utterance.

However, extending UID to referential expressions, particularly those with only two properties, remains largely theoretical. As discussed in Chapter 1.4.1, several assumptions would need to be met. First, it requires treating RER as functionally equivalent to surprisal in UID-based accounts. Second, it assumes that RER correlates with processing cost in a manner comparable to surprisal in sentence comprehension, although empirical support for this link is currently limited (but see Tourtouri et al., 2019 for an initial attempt). Third, most UID evidence derives from broader discourse-level choices or syntactic reductions (e.g., optional use of *that*-complementizer, Jaeger, 2010), rather than fine-grained ordering of content

words in restrictive referential expressions. Moreover, the current project's stimuli, which involve exactly two properties, did not provide a sufficient empirical basis for evaluating UID predictions. The two alternative orders, placing the more or less informative property first, result in symmetrical RER distributions across positions; that is, the RER contrasts (high RER: 2.3 bits vs. low RER: 1.0 bit) are simply reversed, yielding the same mean RER of 1.65 bits. This symmetry renders the overall distribution equally uniform across both structures.

Future research aiming to evaluate UID predictions in referential communication would benefit from designing stimuli that include at least three properties. This would allow for more variation in the distribution of informativity across word positions and enable researchers to assess whether speakers prefer orderings that minimize surprisal peaks to achieve a more uniform distribution of informativity. Additionally, a stronger theoretical grounding would be needed to justify the mapping between RER and surprisal-based processing cost.

Beyond these three aspects, the current project, at a broader level, offers a novel perspective on investigating communication efficiency. Recent theories increasingly converge on the idea that communication efficiency is an influential driving force in shaping human language structure and use (see a review in Gibson et al., 2019; Jaeger & Tily, 2011). Much of the existing evidence comes from corpus analyses, focusing primarily on language form and structure. For instance, word length influences word choice in seeking communication efficiency: Mahowald et al. (2013) found that speakers prefer the shorter word form "chimp" over "chimpanzee" in predictive contexts, while A. F. Frank & Jaeger (2008) observed a similar tendency to reduce the predictable elements in the use of contraction in English (e.g., "couldn't" vs. "could not"). Phonological and acoustic studies have shown that informativity correlates with articulation duration and effort (e.g., Aylett & Turk, 2006; Gahl & Garnsey, 2006; van Son & van Santen, 2005). Syntactically, communication efficiency has been linked to the optional use of "that" complementizer, which helps smooth surprisal peaks in highly informative relative clauses (Jaeger, 2010).

In contrast to these studies, the current project explores communication efficiency from a different angle, focusing on the type of communication involving real-time, visually-situated contexts and collaborative tasks, i.e., the referential communication task (see also e.g., Fukumura, 2018; Tourtouri et al., 2019; Vogels et al., 2020). This type of communication extends beyond the focus on the efficiency of language form and structure, but also addresses other aspects of efficiency, such as the efficiency of interactive communication and efficient allocation of limited cognitive resources (e.g., Hawkins et al., 2021; Lieder & Griffiths, 2020; Simon, 1955, 1972). While these aspects concerning communication efficiency will be further discussed in Section 6.2.3, the central argument here is that the informative-first linearization preference observed in this project provides novel evidence for communication efficiency in the domain of referential production, characterized by its spontaneous, context-dependent, and task-constrained nature.

6.2.2 Reconciling Speaker-Oriented and Listener-Oriented Views: the Mediating Effect of Perspective-Taking

Across five experiments, the current project identified two groups of speakers based on the presence of any syntactic variation in the critical conditions throughout an experiment, while the experiments have already implemented manipulations to encourage and prime the use of both pre-nominal and post-nominal modifications. Speakers were categorized into *Group Consistent*, who consistently used only one syntactic structure throughout the experiment, with most favoring the pre-nominal structure (i.e., action-animal). In contrast, *Group Varied* flexibly employed both syntactic structures, adapting their choices based on informativity.

More interestingly, the proportional distribution of the two speaker groups was not fixed across the five experiments, but was modulated based on the level of communication engagement in different interaction dynamics: the proportion of *Group Varied* significantly increased in Experiments 4 and 5, which involved trial-by-trial role alternations between the Listener and Speaker Tasks for the participants (see Chapter 5 for details), in comparison to Experiments 1-3, where the participants repeatedly performed the same role within an experimental block (Chapter 4). The trial-by-trial role alternation enlarged the proportion of *Group Varied* for two reasons: First, such manipulation introduced the priming effect of the Confederate, potentially making both syntactic structures more active for production during the experiment (a similar effect was observed in Haywood et al., 2003; see also Pickering & Garrod, 2004). Second, such manipulation reinforced frequent perspective-changing between speakers and listeners, inviting the participants to pay more attention to the listeners' visual search processing. A more detailed discussion of this manipulation has been provided in Chapter 5. The present section discusses these results from a more theoretical perspective.

These results contribute to a central question of the current project: whether speakers tailor their referential expressions to listeners' needs (listener-oriented) or rely on heuristic strategies (speaker-oriented). This section first outlines how the two speaker groups, *Group Consistent* and *Group Varied*, align with this theoretical divide between speaker-oriented and listener-oriented views. It then turns to the role of perspective-taking, proposing it as a key mediator in adopting the speaker-oriented or listener-oriented strategies for referential production. Perspective-taking shapes whether speakers engage in audience design, and communication engagement modulates the extent to which speakers would change their perspectives to integrate listeners' perspective during referential production. For speakers who are engaged in perspective-taking, this section further explores how speaker-internal and speaker-external pressures interact during language production, providing support for a parallel processing model (Fukumura, 2018).

6.2.2.1 Evidence for Both Speaker-Oriented and Listener-Oriented Views

The division of the two groups of speakers reflects two diverging theoretical hypotheses on speakers' referential encoding strategies: **The speaker-oriented view** suggests that reference production is primarily heuristic and egocentric (e.g., Horton & Keysar, 1996; Koolen et al., 2013); **The listener-oriented vie**w posits that speakers are engaged in audience design during language production (Clark & Murphy, 1982), aiming to meet listeners' needs, more specifically, to facilitate their visual search for the target in referential communication tasks (e.g., Rubio-Fernández, 2016; Tourtouri et al., 2019).

While mixed evidence and divergent interpretations have been drawn in the line of research focusing on property selection, more specifically, on the high OS rates across different kinds of visual contexts (see Chapter 2.1), the group division repeatedly emerged in the current project focusing on property ordering provided a clearer picture in that both speaker-oriented and listener-oriented speakers can be found in population (similar group division was observed by Tourtouri et al., 2019 in studying the effect of informativity on OS).

As discussed in Chapter 1.4.1 and Chapter 2.3, relying on the preferred word order reflects a speaker-oriented, heuristic strategy of reference production, corresponding to the behavior of *Group Consistent* speakers in the project. This consistent pattern is heuristic because speakers in this group do not have to consider the informativity aspect of visual scenes and therefore do not adjust property ordering accordingly. Instead, they habitually rely on the preferred syntactic structure to describe target figures, regardless of visual contexts or listeners' needs, while still achieving communication success (i.e., describing the target for the listener) to maintain some degree of cooperativeness. Although the current project used a maze-based sentence completion task to increase the difficulty of adhering to this heuristic approach, *Group Consistent* speakers continued to favor it over the listener-oriented strategy in property ordering. Supplementary evidence can be found in the timing dimension of referential production in Experiments 4 and 5, as reported in Appendix A. Speakers in *Consistent Group* began speaking faster on average compared to those in *Group Varied*, both in the Maze Task and during spoken production. This suggests that adhering to the preferred structure requires less speaking effort, supporting the notion that this behavior is heuristic.

By contrast, a listener-oriented strategy for property ordering requires speakers to take the needs of listeners into account. This approach goes beyond achieving communication success at the surface level, but requires speakers to actively pursue communication efficiency by employing informativity in property ordering to enable efficient target search for listeners. This listener-oriented approach corresponds to the behavior of *Group Varied* speakers, who are sensitive to informativity and use it to arrange properties in an order that efficiently supports the listener-oriented goal.

This encoding strategy based on informativity has also been observed in property selection, where OS rates vary as a function of RER (Tourtouri et al., 2019), as well as in adjective ordering, where the more discriminating property tends to be mentioned first (Fukumura, 2018). The current project extends these findings by demonstrating that informativity also influences property ordering at the syntactic level, as reflected in speakers' use of pre-nominal versus post-nominal modifications.

Importantly, this result, together with the timing evidence reported in Appendix A, offers stronger support for distinguishing listener-oriented strategies from speaker-oriented heuristics than previous research on OS alone (see Chapter 2.2.2.5 and Chapter 2.1.3.1). This is because, although producing post-nominal modifications entails greater production effort than pre-nominal ones (see also Brown-Schmidt & Konopka, 2008), *Group Varied* speakers would expend this increased cost to produce post-nominal modifications with a listener-oriented communicative goal, highlighting their consideration of communication efficiency over production ease during the referential communication task.

6.2.2.2 The Mediating Effect of Perspective-Taking

The current project not only confirms the existence of both speaker types, as already suggested by Tourtouri et al. (2019), but also identifies factors that influence the proportional distribution of the two groups. One key factor is perspective-changing, which was enhanced in the current experiments through trial-by-trial role alternation. This role alternation encouraged more speakers to adopt a listener-oriented strategy by shifting their attention toward the listener's perspective.

In this project, when participants performed the listener role, they were consistently exposed to expressions following the informative-first linearization preference, as the confederate was designed to be always informative. This manipulation not only emphasized the efficiency aspect of the communication system (Shannon, 1948), but also maximized participants' direct experience of efficient visual search guided by the informative-first ordering. As a result, speakers were more likely to adopt a similar linearization pattern when producing their own referential expressions for listeners.

Another factor, also linked to perspective-changing, is the status of the communication partner. The "status" can refer to whether the partner's role was actively present for speakers. In Experiments 1- 3, the simulated Confederate, who only provided textual feedback within blocks, was perceived as less involved compared to Experiment 4, where participants believed their partner was actively performing the same task in turn. The "status" can also refer to the authenticity of the partner's role — whether the partner was perceived as the computer machine, a virtual agent, or another human participant, as compared in the online communication game in Experiment 4 and the face-to-face communication in Experiment 5 (see also Peña et al., 2023). Another type of "status" was tested by Haywood et al. (2003), which is the perceived "helpfulness" of a confederate (see more relevant factors in Chapter 2.2.2.4). These factors create different communication dynamics, contributing to speakers' sense of communication engagement, making them more likely to consider the listener's perspective in referential production.

As discussed in Chapter 2.2.2.4, while perspective-taking has been shown to influence speakers' non-restrictive utterances, ongoing debates concern whether and to what extent perspective-taking, and so audience design, is involved in producing utterances that are more restricted by stimuli, such as conventional targets adopted in referential communication tasks. Similar debates, introduced in Chapter 1.3.2, revolve around the influence of speaker-internal versus speaker-external factors on language production.

Previous studies adopting the director-matcher paradigm, where speakers are exposed to privileged knowledge, strongly argue that speakers primarily adopt an egocentric perspective in referential production. They consider listeners' perspective and engage in audience design only to a limited extent and in the later stage of language production (e.g., Damen et al., 2019; Horton & Keysar, 1996; Keysar et al., 1998; Wardlow Lane et al., 2006). This is evidenced by speakers frequently including privileged knowledge in their referential expressions, thereby failing to exclude irrelevant information for listeners. Speakers also do not always avoid ambiguity that may confuse listeners (e.g., Arnold et al., 2004; Ferreira et al., 2005; Ferreira, 2008; Ferreira & Dell, 2000). The main motivation for this claim is that taking the perspective of others requires intensive cognitive resources (e.g., Keysar et al., 1998). By contrast, other studies suggest that speakers incorporate listeners' perspective during the very early grammatical encoding stages that influence syntactic choices (e.g., Lockridge & Brennan, 2002, see Chapter 1.3 and Chapter 2.2.2.4 for more details).

The current project contributes to this debate by providing insights into how perspective-taking influences speakers' referential encoding strategies. It identifies both informativity and perspective-taking as speaker-external, communication-driven factors, while examining property ordering, rendering syntactic variation, can reflect effects at the grammatical encoding stage of language production, specifically, during positional processing, an early stage of utterance planning according to Levelt (1989)'s language production model (see Figure 1.3 in Chapter 1.3.1). Addressing this question would also deepen the understanding of the debate between speaker-oriented and listener-oriented views.

The proposal here is to consider perspective-taking as a mediating factor that links the effect of audience design with reference production in referential communication tasks.[1] Whether speakers adopt a speaker-oriented or listener-oriented strategy in referential encoding should largely depend on the extent to which they engage in perspective-taking during communication. At the same time, perspective-taking itself is a flexible process (see General Discussion in Chapter 5, see also Hawkins et al., 2021), modulated by factors across multiple dimensions, including individual differences such as executive function and working memory (Brown-Schmidt, 2009; Ryskin et al., 2015; Wardlow Lane, 2013), communication tasks and goals (Vogels et al., 2020; Yoon et al., 2012), social skills (e.g., Mendelsohn & Straker, 1999), and cross-cultural differences (S. Wu & Keysar, 2007), etc.

The current project highlights one of the aspects, which is communication engagement. This is evidenced by the increased proportion of *Group Varied* speakers in Experiments 4 and 5, where the chance of perspective-changing was increased by trial-by-trial role alternations and increased authenticity of the communication partner.

[1]Notably, the argument about the role of perspective-taking here primarily addresses the debate between the speaker-oriented and listener-oriented views in the topic of property selection and ordering, rather than studies using the director-matcher paradigm, due to fundamental differences in experimental design. For the role of perspective-taking in the director-matcher paradigm, see e.g., Horton & Keysar (1996) and Keysar et al. (2000).

Involving the Listener's Perspective Speakers are more likely to adopt a listener-oriented approach when they are more engaged with the listener's perspective during communication. For those who are indeed engaged in perspective-changing, they would plan the utterances by incorporating listeners' perspectives. Such planning guided by informativity and perspective-changing can appear early in language production: results of the current project indicate that informativity affects property ordering at the syntactic level. According to Levelt (1989)'s model, this decision must occur early during the grammatical encoding stage, as the relative positions of properties, as well as the functional assignment of lexemes to lemmas (e.g., assigning the action property CRY as an adjective "crying" or as a relative clause "that cries" in German), are determined along with the syntactic structure.

However, the influence of this speaker-external factor (i.e., informativity) may still compete with speaker-internal factors within individual speakers. This is suggested by the observation that although the overall proportion of *Group Varied* speakers increased across experiments, particularly in more engaging setups like Experiment 5, the effect size of the informativity within *Group Varied* remained relatively stable. Even when a greater number of speakers adopted an informativity-first linearization strategy, the degree to which informativity modulated their syntactic choices did not increase. On average, the difference in post-nominal modification rates between the Animal-Informative and Action-Informative Conditions among *Group Varied* speakers remained around 5% across the five experiments, indicating a modest but consistent effect size. This suggests that while communication engagement may increase the likelihood of adopting the listener-oriented strategy for property ordering, it does not necessarily amplify the effect strength of informativity within individual speakers.

This is probably because even among *Group Varied* speakers, there remains a persistent bias toward the pre-nominal structure. If word order preferences are considered as a speaker-internal factor, this suggests that such preferences also exert an influence overall, shaping referential choices even when speakers take listeners' perspective into account. Another speaker-internal factor is priming, which also influences property ordering in the current project, independent of the effect of informativity (see Chapter 5 for details).

The co-presence of speaker-internal and speaker-external effects in the current project supports the Parallel-Activation-for-Selection-and-Sequencing (PASS) model proposed by Fukumura (2018). This model posits that the effects of informativity and availability are activated in parallel and *simultaneously* influence property ordering decisions during the planning of referential expressions. In Fukumura (2018), speakers described color and pattern properties of objects, and their choice of adjective order reflected both the availability effect of color property and the informativity of each property, as determined by discriminability. The results showed that color adjectives were often produced first when they were more available and/or more informative, but when pattern adjectives were more informative, speakers were also more likely to place the pattern adjective first and invert the preferred color-pattern adjective order (see Figure 2.9 in Chapter 2.2.2.3). The joint effects of informativity and availability support the PASS model, where both effects are active simultaneously in influencing adjective ordering. This model

is in contrast with other incremental models (e.g., Dale & Reiter, 1995) that propose a serial processing algorithm, where different factors determine property selection step by step, with availability taking precedence over informativity.

The findings of the current project further support the PASS model, particularly if the concept of availability can be extended to encompass other speaker-internal factors such as word order preferences and priming. Specifically, the results suggest that informativity, as a speaker-external factor, operates in parallel with these speaker-internal factors when determining syntactic linearization among speakers engaged in perspective-taking, i.e., among *Group Varied*. Notably, the relative strength of these effects appears stable within individual speakers in *Group Varied*, as indicated by the similar proportional distribution of the two syntactic structures in the two critical conditions across experiments (i.e., a similar effect size of RER). This suggests that the influences of informativity, priming, and word order preferences are not mutually exclusive or overriding one another in determining property ordering, but operate jointly and concurrently in shaping speakers' syntactic choices in the current project.

This pattern stands in contrast to the hypothesis that posit a strictly serial processing mechanism, where speaker-external factors and the listener's perspective are integrated only at later stages of production planning (e.g., Horton & Keysar, 1996; Keysar et al., 1998). Instead, the current findings align with the view that speaker-external factors can influence reference production at early stages of grammatical encoding, consistent with the PASS model.

Taking the Speaker's Perspective In contrast, not all speakers showed evidence of integrating the listener's perspective into their syntactic choices. When speakers are not engaged in perspective-taking extensively, they are more likely to adopt speaker-oriented, heuristic approaches. The extent to which speakers are engaged in perspective-taking heavily depends on the experimental manipulations. Especially when the manipulations are more moderate, such as block design in Experiments 2 and 3 and in previous research, such as Vogels et al. (2020), speakers may choose not to change perspectives towards listeners, or at least not to apply it in referential encoding strategies. Instead, they may rely primarily on their own perspective, adopting a speaker-oriented, heuristic approach. This is evidenced by the repeated presence of *Group Consistent* speakers in the current project, as well as by a group of speakers identified in Tourtouri et al. (2019), who universally overspecified color and pattern adjectives regardless of the RER dimension of visual contexts.

Similarly, studies on OS, such as Gatt et al. (2013) and Koolen et al. (2011), where participants were not explicitly required to shift perspectives, also indicate that when perspective-taking is not strongly cued, speakers are more likely to focus on their own perspective and adopt a speaker-oriented, heuristic approach in referential production. The speaker-oriented approach would be even more likely to be adopted if speakers' own perspective is further highlighted by task demands, such as the informing task adopted in Yoon et al. (2012), or when speaker-internal factors were more prominent, such as boosting salience of certain visual properties in Wardlow Lane & Ferreira (2008). The motivation for this heuristic approach

will be further discussed next in Section 6.2.3 relating to production efficiency and resource allocation.

Taken together, the current project shows evidence that both types of speakers, adopting either the speaker-oriented or the listener-oriented referential encoding strategies, can be found in the population. Furthermore, the debate between the two views may be reconciled by the flexible perspective-taking process as the mediating step to link speakers' referential encoding and audience design. Perspective-taking can be affected by the level of communication engagement of the tasks and communication contexts. Speakers are more likely to focus on their own perspectives if the communication environment does not provide active involvement of the addressee or if the task focuses more on the speaker's perspective. They would then prefer the speaker-oriented, heuristic approach for easy production (e.g., Goldberg & Ferreira, 2022; MacDonald, 2013). By contrast, speakers are more likely to attune to the listener's perspective when the communication task is interacting and engaging, as it emphasizes the addressee's perspective and joint communication goals. In such contexts, audience design is more likely to be employed, operating simultaneously with speaker-internal factors, in informative property ordering, ultimately aiming for communication efficiency.

6.2.3 Revisiting the Notion of Communication Efficiency

As the final discussion point of the current project, it is important to revisit the notion of *communication efficiency* in referential communication tasks. This project defines efficiency primarily from the listener's perspective, assuming that an efficient communication process — and, by extension, an informative referential expression — is one that facilitates target identification, which is a listener-oriented goal, because in this task, it is the listener who ultimately completes the communication goal, by e.g., clicking on the target.

Previous research has presented various perspectives and corresponding measures regarding what constitutes an *efficient* visual search process for listeners. For instance, a series of eye-tracking studies (e.g., Rubio-Fernández et al., 2021; Rubio-Fernández & Jara-Ettinger, 2020) closely links efficiency to incrementality, arguing that target visual search should align with the sequential presentation of properties (see Section 2.2.2.1 and Section 2.2.2.5 in Chapter 2.1). Incrementality is then related to both the timing and the number of fixations on the target, with earlier and more frequent fixations indicating a more efficient visual search process, because this suggests that listeners start searching for the target as early as possible guided by the time-locked utterance of property words that progressively narrow the selection scope (e.g., Eberhard et al., 1995).

Apart from the timing and proportion of fixations, Tourtouri et al. (2019) links efficiency to overall lower cognitive effort required during visual processing, using the Index of Cognitive Activity (ICA) as a measure of cognitive load (see Chapter 2.1.2.1). Specifically, they correlate RER with ICA, showing that informative words (high RER) demand greater processing effort during comprehension. An informative OS expression, such as "the *blue* dotted ball," where "blue" is overspecified but informative (see Condition C in Figure 2.4), can enhance efficiency by reducing

the processing effort appearing later at the head noun. This is evidenced by the lower ICA values at the head noun, e.g., at the word "ball", the point where the target is finally identified. Other measures of efficiency for comprehension include listeners' reaction times of clicking on the target (e.g., Fukumura & Carminati, 2021; Tourtouri et al., 2019) and the N400 ERP component (e.g., Engelhardt et al., 2011). While these studies yield some detailed incongruity in their findings (see Chapter 2.1.2.1 for an overview), they generally support the idea that a more efficient, facilitated visual search process is the one that enables listeners to identify targets more quickly with reduced processing effort.

Defining efficiency from the speaker's perspective, on the other hand, is more debatable. From a listener-oriented view, an efficient referential encoding should align with the notion of efficiency from the listener's perspective. That is, the content and structure of an efficient referential expression should facilitate faster target identification and/or reduce processing effort for listeners. As Rubio-Fernández (2016) argues, from the speaker's perspective, the key issue is not whether listeners experience an facilitation effect (see Chapter 1.3.2.1), but whether speakers act with a cooperative intention, that is, whether they reasonably expect how listeners will process and benefit from speakers' utterances. If speaker's referential expressions reflect this cooperative intention to support the listener, they can be considered listener-oriented and efficient. The behavior of *Group Varied* speakers in the current project reflected this listener-oriented intention of efficiency, as they adjusted property ordering according to informativity to facilitate the listener's comprehension in referential production.

From the speaker-oriented view, however, if mirroring the definition of efficiency for comprehension, production efficiency could be defined in terms of initiating speaking as early as possible with minimal production effort. Referential encodings would be efficient for production if speakers begin speaking as soon as possible, with a tendency to first utter a construction that is already available (e.g., Bock & Irwin, 1980; Ferreira, 2008; MacDonald, 2013). This typically includes properties that are visually salient and easy to name, structured in an order and syntactic form that speakers habitually prefer. From this perspective, the speaker-oriented, heuristic approach should also be seen as an efficient approach, which also emerged repeatedly in the current project as the *Group Consistent* speakers.

Notably, these two perspectives of efficiency may not always align in their predictions of referential encoding. For property selection (and omission), the predictions of the two perspectives can align, as higher OS rates may reflect not only an informative, listener-oriented encoding strategy (see Chapter 2.1), but also a production advantage: including an extra property, especially prenominally, provides additional time for planning the upcoming utterance, due to the inherent nature of incrementality of language production (e.g., Pechmann, 1989). This line of reasoning also applies to the optional use of "that" complementizer in support of the UID hypothesis (Jaeger, 2010). Apart from seeking uniform distribution, including "that" also prolong time for planning the upcoming relative clause (see also e.g., Ferreira, 2008; Ferreira & Dell, 2000), especially when the beginning of the relative clause contains high surprisal values, which often correlates with lower availability and greater production difficulty due to e.g., low word

frequency. The inclusion of optional words seems to facilitate both comprehension and production efficiency.

In property ordering, by contrast, the two types of efficiency may oppose and compete, resulting in divergent, rather than aligned, predictions that would jointly shape the distribution of speakers' linearization preferences. The most available word or structure for speaking may not necessarily be the most informative and efficient form for listeners' comprehension. This tension is evident not only in the two speaker groups identified in the current project but also in Fukumura (2018) and Haywood et al. (2003), where both informativity and availability were found to influence property ordering (see above in Section 6.2.2.2).

Beyond referential communication tasks, the tension between comprehension- and production-based efficiency remains. A classic example comes from Ferreira & Dell (2000), who investigated whether the optional inclusion of "that" in object clauses (e.g., "You/I knew (that) you/I missed the practice") is primarily driven by availability or by a listener-oriented need to prevent ambiguity. The availability-based account predicts that "that" is omitted more often when the subject pronoun in the embedded clause repeats the subject pronoun of the main clause (i.e., "you knew you" or "I knew I"), as the repetition increases the availability of the pronoun. In contrast, the listener-oriented account argues that "that" should be retained to avoid ambiguity, as in "I knew *you*...", where the embedded subject "you" holds temporary ambiguity without including "that", versus the unambiguous condition "You knew *I*...", where "I" has to be the subject of an object clause instead of an object. Across six experiments, Ferreira & Dell (2000) consistently found that availability played a dominant role, i.e., "that" was more likely to be omitted when the subject pronoun was repeated, while there was little evidence supporting ambiguity avoidance, even in communicative tasks, as the frequency of "that" use remained to be similar regardless of ambiguity when pronouns differed.

However, research in information theory, such as Pate & Goldwater (2015), criticized that the availability-based efficiency in Ferreira & Dell (2000) should not be considered a type of "communication efficiency" but rather only as "short-term ease" for speakers. They argued that ambiguity avoidance represents only a categorical aspect of listener-oriented needs. More broadly, omitting "that" may increase the risk of communication errors in a noisy channel, where additional linguistic cues help ensure accurate message transmission.

A more moderate hypothesis reconciles these opposing forces by proposing a trade-off between speaker and listener effort in communication: if speakers minimize effort in constructing their utterances, listeners must invest more effort to interpret them, and vice versa (termed as "division of labor" in Ferreira, 2008). Under this view, communication efficiency is defined as minimizing the overall effort within the system, predicting that an efficient system should minimize the average of production and comprehension effort between speakers and listeners (e.g., "minimizing collaborative effort" by Clark & Wilkes-Gibbs, 1986; Gibson et al., 2019).

Evidence from multimodal communication, where gestures accompany speech, supports this trade-off. Examining repair initiation (e.g., asking "which one?" by the speaker) and repair resolution (responding by the listener), Rasenberg et al. (2022) systematically identified a trade-off in labor division between the

interlocutors, in both speech and gesture effort. Moreover, they found that speakers tend to prefer to initiate "restricted offers" in repair (e.g., "like this?") over open-ended questions (e.g., "huh?" or "who?"). They quantitatively showed that the restricted offers cost minimal "joint multimodal effort" within the communication system, possibly because offering concrete and restricted options is more informative to solve the repair. An interesting direction for future research in referential communication tasks is to develop experimental designs and measurements that can systematically capture the *joint effort* of a communication unit, rather than focusing solely on production or comprehension in isolation.

Another aspect of efficiency concerns the allocation of cognitive resources. Speakers in referential communication tasks engage in multiple cognitive processes, including visual processing, language production, and potentially perspective-taking, requiring them to strategically distribute their limited cognitive resources while multitasking. Hawkins et al. (2021) and Lieder & Griffiths (2020) argued that the allocation of cognitive resources is rational and efficient, proposing *the resource-rational model*, which builds on the notion of bounded rationality (Simon, 1972). Specifically focusing on resource allocation for perspective-taking in referential communication, Hawkins et al. (2021) suggested that the extent to which speakers engage in perspective-taking depends on a trade-off between its cognitive cost and communicative benefit. The two groups of speakers observed in the current project may provide evidence for the resource-rational model in referential production, as some speakers appeared to reduce cognitive cost by not engaging extensively in perspective-taking when cognitive demands are high for them (i.e., *Group Consistent*), while other speakers emphasize more on communicative benefit and choose to expand more cognitive resources for perspective-taking (i.e., *Group Varied*). More generally, Lieder & Griffiths (2020) emphasized that heuristic approaches should not be dismissed as "irrational" but rather understood as rational strategies for allocating limited cognitive resources.

Discussion in this section intends to highlight the need for a more precise and comprehensive definition of "communication efficiency" in research on language production and interactive communication. While information-theoretic accounts and psycholinguistic studies supporting audience design emphasize efficiency from a listener-oriented perspective, the persistent influence of availability-based, speaker-oriented heuristic approaches for efficiency in language production should not be overlooked.

6.3 Limitations and Future Directions

This final section addresses the limitations of the current project, primarily the artificial constraints of the referential communication task. It will then propose three directions for future research: 1) examining the facilitation effect of word order in comprehension, 2) extending the effect of informativity to other structures with ordering alternations, and 3) exploring human-to-computer communication.

6.3.1 Limitations of the Referential Communication Task

The major limitation of the current work stems from the nature of the referential communication task itself. Critiques include that the task is constrained and artificial. Rubio-Fernández (2017) questioned the extent to which findings from such tasks can be generalized, particularly targeting the director-matcher paradigm with occluded objects. Similar concerns may also apply to referential communication tasks more broadly. One of her major critiques is that the setup of this task is artificial, as interlocutors' referential domain is highly restricted to the visual scene presented, and so their utterances do not fully reflect the complexity and fluidity of real-life referential interactions (see also Bezuidenhout, 2013; Rubio-Fernández et al., 2024).

Additionally, referential communication tasks impose high and somewhat artificial cognitive demands. For example, perspective-taking requires inhibitory control to suppress one's own perspective in favor of the listener's (e.g., Nilsen & Graham, 2009). Similarly, producing the less preferred post-nominal structure may necessitate inhibiting the preferred syntactic preferences. These cognitive demands may require speakers to expend more cognitive effort in experimental settings than in natural referential interactions, where such skills are engaged more flexibly. Thus, failing to meet task expectations does not necessarily indicate a real lack of pragmatic reasoning or perspective-taking skills of speakers.

Another limitation is that referential communication tasks often involve restricted utterances, meaning that the message encoding phase in production planning is fixed by the target referents. This differs from the flexibility of incremental language production in daily communication. Hawkins et al. (2021) compared the informativity of speakers' scripted and unscripted utterances, and they argued that speakers' encoding of referents may reflect an adaptation to the informativity expectations imposed by the task itself. They found that speakers' informativity levels naturally increased when they were allowed to produce unscripted utterances[2], suggesting that experimentally imposed informativity expectations may not align with spontaneous, unrestricted utterances. In the current project, this may have influenced participants' adaptation to the Confederate's behavior: since the Confederate consistently used an informative-first linearization preference, participants may have adjusted their expectations and linearization strategies accordingly. Furthermore, allowing unscripted utterances (e.g., "the rabbit that is painting on a board with a brush") would introduce additional factors influencing syntactic choices, such as the length and complexity of relative clauses.

Several studies have developed innovative experimental paradigms that reduce these task-imposed constraints while maintaining careful experimental control. For example, Gussow & MacDonald (2023) investigated cases where speakers begin speaking without knowing the full message in advance. They presented objects in two pairs (e.g., "the vest and the pear" as one pair), while the target pair, which

[2]Notably, Hawkins et al. (2021) defined informativity as the extent to which an utterance helps disambiguate the target (e.g., referring to "a cassette tape" as "the tape" would be uninformative if "a roll of tape" were a competitor). This definition may influence the conclusions and interpretations drawn from their study, as it inherently favors the idea that using more words contributes to greater disambiguation.

speakers had to describe, appeared later. One of the objects in the target pair either overlapped with an object in the other pair (e.g., "the vest and the chair" as an overlap condition) or did not. They found that speakers strongly preferred to mention the overlapping object first, suggesting that when speakers do not have access to the full message beforehand, they favor early and incremental utterance planning by uttering first the more available object.

Koranda et al. (2022) introduced a paradigm using pseudowords to describe unconventional compass directions (e.g., "blit" for a direction of 105 degrees left to the north). These pseudowords were presented with varying frequency in the word-learning phase. During the testing phase, speakers guided a treasure hunter toward a direction that was either near or far from those learned directions. Although low-frequency words were more precise for the addressee, speakers overwhelmingly preferred high-frequency words, which were more available due to frequent exposure in the learning phase.

While these studies unsurprisingly support an availability-based account of language production, future research could extend these novel paradigms to address the artificiality of referential communication tasks. Building on Gussow & MacDonald (2023), where speakers cannot fully encode a message before speaking, future studies could involve visually complex objects requiring syntactic structures such as pre- or post-nominal modification (e.g., "the blue vest and the chair that is red" if color would affect the availability and informativity in object pairs). Adapting the pseudoword paradigm by Koranda et al. (2022), researchers could infer informativity based on participants' use and expectations of pseudowords during an experiment, rather than setting a fixed linguistic standard based on conventional stimuli.

6.3.2 Future Directions

6.3.2.1 How Exactly does Informative-First Linearization Preference Facilitate Target Visual Search?

An immediate future direction following the current project should investigate comprehension processing of listeners, to determine whether and how exactly the informative-first linearization preference facilitates listeners' target identification. While previous cross-linguistic research has shown that listeners search for targets incrementally (Rubio-Fernández et al., 2021; Rubio-Fernández & Jara-Ettinger, 2020), further research is invited to examine the comprehension of utterances involving multiple property ordering options and syntactic variations.

A potential issue in previous cross-linguistic research (e.g., Rubio-Fernández et al., 2021; Rubio-Fernández & Jara-Ettinger, 2020) is that although both word orders were included, each was the preferred word order for the respective native speaker group. This between-subjects design may yield qualitatively different conclusions compared to within-subjects designs that directly contrast preferred and less preferred word orders in individual speakers' visual search behavior. An important question remains: to what extent can a less preferred but highly informative word order facilitate visual search? This raises a potential competition between

the facilitation effect of visual search guided by informativity and the processing difficulty associated with comprehending the dispreferred property ordering.

Previous studies (mainly, Huang & Federmeier, 2012; Kemmerer et al., 2007) have used ERP measures to demonstrate that reversed adjective ordering (i.e., the less preferred order of color-size adjectives, such as "the grey huge elephant") elicits processing difficulties for comprehension. Using an acceptability rating task, Kemmerer et al. (2007) first identified two groups of comprehenders with high versus low sensitivity to the less preferred adjective ordering. While their primary focus was on the high-sensitivity group, they also showed that these groups exhibited different ERP patterns across word positions. For the high-sensitivity group, processing difficulties were primarily observed at the second adjective, where they detected a reduced N400 and an enlarged P600 in reversed order (e.g., at "huge" in "the grey huge elephant") compared with the preferred order, suggesting both semantic and syntactic reanalysis of the ordering constructions. However, Huang & Federmeier (2012) argued that reversed ordering may primarily impact semantic, rather than syntactic, processing difficulty, as their study found increased N400 effects at both the second adjective and the final noun position.

Regardless of the detailed divergent results of ERPs, these studies primarily focused on the comprehension of the linguistic forms of adjective ordering, but not involving how they interact with visual search. In addition, the reversed ordering does not involve syntactic alternations (see also Fukumura & Carminati, 2021). A potential hypothesis would be that the low-sensitivity group, detected in Kemmerer et al. (2007), may find the informative-first order more beneficial during visual search than the high-sensitivity group, who may experience more comprehension difficulty of the reversed order.

Another concern regarding previous studies is that the experimental conditions did not allow for an equal gradient reduction of uncertainty. For instance, in Rubio-Fernández et al. (2021), the target referent "a blue triangle" had only one competitor in terms of color ("a red triangle") or shape ("a blue star") across two conditions. This design inherently led to full certainty about the target based on the very first property in one of the conditions: e.g., "blue" immediately disambiguates the target referent for speakers of pre-nominal languages, while "triangle" does so for post-nominal speakers. Although such a design can be seen as an extreme case of entropy reduction, it is less convincing that early fixations towards the target truly reflect a facilitation effect of word order or simply an artifact of the visual display structure, which allowed participants to pinpoint the target as soon as the first property was mentioned.

The need for further research on the effect of word order on comprehension is also supported by findings from Fukumura & Carminati (2021), which indicate that the facilitation effect is property-specific, where color adjectives contribute more to efficient target identification than pattern adjectives in cases of OS. Similarly, Tourtouri et al. (2019) found overall lower ICA values for color adjectives than for pattern adjectives, suggesting that color information is generally easier to process. To generalize the conclusions drawn by studies such as Rubio-Fernández et al. (2021) and Rubio-Fernández & Jara-Ettinger (2020) on word order facilitation, it is crucial to explore whether these effects can be found in other visual properties

and ordering variations, such as post-nominal modifications and the less preferred adjective orders. The stimuli and ordering structures used in the current project are already well-suited for extending this line of research in language comprehension.

6.3.2.2 Effects of Informativity on Other Syntactic Structures

Another important direction is to extend the effect of informativity on property selection and ordering to other linguistic structures with meaning-equivalent alternatives. A classic example is the distinction between Double Object (DO) constructions (e.g., "Give me the letter") and Prepositional Object (PO) constructions (e.g., "Give the letter to me") (e.g., Pickering et al., 2002). Arnold et al. (2004) examined local PP-attachment ambiguity, where attaching an additional prepositional phrase in PO constructions may introduce ambiguity for listeners (e.g., "Give the letter *to Kim* to me"). In two communicative experiments, an addressee asked speakers to recall and rephrase a stimulus sentence (e.g., asking "What did the foundation do?" after speakers read "A museum received Grant's letters to Lincoln from the foundation"). Speakers could respond with either an unambiguous DO structure (e.g., "The foundation sent a museum Grant's letters to Lincoln") or an ambiguous PO structure with PP attachment (e.g., "The foundation sent Grant's letters *to Lincoln* to a museum"). The results showed that speakers neither avoided producing the ambiguous PP-attachment structures nor employed prosody for disambiguation. Instead, structural choices were primarily driven by speaker-internal factors, such as verb-specific structural biases and the length of each object constituent, rather than audience design. Similar conclusions were drawn by Morgan & Ferreira (2022), who investigated the production of resumptive pronouns in relative clauses with long dependencies (e.g., "It was the little boy who we wondered why someone hit _____." It can be filled with either a resumptive pronoun *him* or left as it is without continuation). The use of the resumptive pronouns was not affected by the listener's level of acceptability of resumptive pronouns in complex sentence structures. Morgan & Ferreira (2022) also points out that there is so far little evidence in "syntactic audience design" in research focusing on language production processing.

Although the *Group Varied* speakers identified in the current project provide some evidence in support of syntactic audience design, with a small but stable effect size, the complexity of pre- and post-nominal structures is by no means comparable to Arnold et al. (2004) and Morgan & Ferreira (2022), who examined the production of more complex syntactic structures. Although null effect of audience design has emerged repeatedly for complex structures, future research may also continue to explore the role of speaker-external factors beyond categorical factors in audience design (Pate & Goldwater, 2015), such as ambiguity avoidance, but seeking quantifiable measures of audience design by e.g., varying the surprisal (i.e., informativity) of linguistic components, such as the two object constituents in DO and PO constructions.

Another promising avenue for investigating the role of informativity in language production is the optional mention of instruments in sentences depicting events. A converging observation is that speakers tend to mention atypical instruments more frequently than typical ones (e.g., "stabbing with an ice pick/a knife," Brown

& Dell, 1987), especially for naive addressees (Lockridge & Brennan, 2002, as discussed in Chapter 2.2.2.4). Comprehension studies by Rohde et al. (2021) and Rohde et al. (2022) have demonstrated that the processing difficulty of atypical instruments for comprehenders is mediated by expectations established from the contexts, specifically by factors related to surprisal and informativity (e.g., "Mary is a boring/surprising person. She chopped carrots with a shovel." Example adapted from Rohde et al., 2021, where "surprising" reduced the reading time of the atypical instrument "shovel").

The optional mention of instruments in language production remains an under-explored area, making it an intriguing case for future research. First, encoding an instrument is not obligatory, as a typical instrument is often omitted, making it more likely to be sensitive to information-theoretic factors compared to expressions with stricter grammatical constraints, such as pre-nominal modifiers. In addition, at least in English, instruments are conventionally encoded at the end of a sentence, which avoids confounding with the incremental, availability-based production strategies, i.e., speakers cannot rely on it to buy time for future planning (see also Section 6.2.3). Finally, instruments can be alternatively ordered in a sentence using a marked structure (e.g., "She used a shovel to chop carrots"). If the use of this alternative ordering can be modulated by the typicality of instruments, it would provide valuable insights into the influence of informativity on linearization and syntactic audience design.

6.3.2.3 Human-to-Computer Communication

As briefly discussed in Chapter 5, the experiments designed in the current project align well with research on *Human-to-Computer Communication (HCC)*. This field is rapidly evolving with advancements in speech technology that support various communication modalities (e.g., Aylett et al., 2014; Bernard & Arnold, 2019). A key question in HCC research is how humans perceive and respond to computers as interlocutors. The influential Computers Are Social Actors (CASA) framework hypothesizes that people intuitively interact with computers as if they were human partners (e.g., Nass & Moon, 2000). A crucial test of this hypothesis, relevant to the current project, is whether and to what extent pragmatic and communicative principles from *Human-to-Human Communication (HHC)* also apply to HCC, particularly in topics such as cooperation, common ground, and audience design.

One approach to this question has been to examine the alignment effect in HCC. Alignment, the tendency to repeat interlocutors' linguistic constructions, has been interpreted as evidence of both automatic priming (e.g., Pickering & Garrod, 2004, see also Chapter 2.2.2.2 and Chapter 5) and audience design, with speakers adapting their language use with other interlocutors as a way to facilitate communication and social rapport (e.g., Brennan & Clark, 1996; Chartrand & van Baaren, 2009; Ferreira, 2019; Shen & Wang, 2023). If speakers indeed perceive and interact with a computer as a social entity, alignment should also be observed in HCC.

Alignment in HCC has been demonstrated at both the lexical (Branigan et al., 2011; Shen & Wang, 2023) and syntactic levels (Cowan et al., 2015), with speakers showing an increased tendency to use the less preferred linguistic constructions (e.g.,

less frequent lexicons, post-nominal modifications, and PO structures). Especially for lexical alignment, Branigan et al. (2011) and Shen & Wang (2023) showed that lexical alignment appears to be even stronger in HCC than in HHC: speakers tend to maintain lexical alignment more rigidly when interacting with computers, evidenced by a larger proportional use of the less preferred, but primed lexicons. Research of HCC alignment thus argues that human speakers may involve stronger audience design by considering the status of the addressee (i.e., the computer) as less competent in communication than human interlocutors, and thus align with the computer to a greater extent in pursuing communication success (see also Rothwell et al., 2021).

However, research using the director-matcher task has suggested that audience design in HCC may not be as prominent as suggested in research on alignment. Peña et al. (2023), discussed in Chapter 5, found that speakers were more likely to include privileged knowledge when addressing a computer than when speaking to a human partner. Notably, this effect was modulated by the perceived competence of the computer: when speakers perceived the computer as an advanced dialogue agent rather than a simple computer, they treated it more like a human interlocutor and included privileged information less frequently, achieving an equal, but not higher, level of audience design with the computer agent, evidenced by a similar proportion of privilege knowledge inclusion between HHC and HCC.

Although the current project does not primarily focus on HCC, it contributes to this field by comparing HHC and HCC using the referential communication task. First, it replicates the syntactic priming effect reported by Cowan et al. (2015) in Experiment 4, where speakers aligned their structural choices with the simulated Confederate producing synthesized, human-like speech, suggesting some level of alignment in structural use between human speakers and the computer partner. However, whether this alignment tendency can be further interpreted a form of audience design is beyond the research scope of the current project. The priming effect observed in Chapter 5 was immediate, occurring within a single trial turn, which may reflect more automatic priming (Pickering & Garrod, 2004) rather than the alignment reflecting audience design (e.g., Branigan et al., 2011; Cowan & Branigan, 2015; Shen & Wang, 2023). Studies supporting alignment as audience design often employ more complex design in trial orders, such as inserting filler trials between the priming and experimental trials, which was not implemented in the current study. Therefore, the observed priming effect in the current work may not allow for a clear distinction between automatic priming and audience design.

Second, the current project challenges the CASA framework. Previous studies comparing HCC and HHC often implemented the HHC condition by using human-recorded audio files (e.g., Peña et al., 2023; Shen & Wang, 2023). However, the experiment format remained computer-mediated (i.e., participants interacting with a computer at home or in the lab). These studies frequently conclude that audience design in HCC is as strong as, or even stronger than, in HHC. The current project challenges this claim by demonstrating that when HHC occurs in a more natural, in-lab, face-to-face setting with higher communication engagement, speakers' performance is, in fact, more listener-oriented than in HCC. This is evidenced by the highest proportion of *Group Varied* speakers in Experiment

5, conducted as an in-person HHC, compared to Experiments 1- 4, which were conducted online in the HCC format.

Future research should examine the CASA framework more carefully when comparing HCC and HHC. Especially given that non-verbal communicative cues such as disfluency (e.g., Oviatt, 1995) and gesture (e.g., Mol et al., 2009) are largely absent in HCC, the hypotheses proposed by CASA should be further tested to determine when and in which communication dimensions they are applicable. This can be achieved by exploring other communicative elements or experimental tasks that are well-established in HHC.

6.4 Conclusion

This thesis examined the effect of informativity on syntactic linearization of property ordering in referential production, focusing on whether speakers adjust property ordering based on the listener's need of efficient target identification. By integrating insights from psycholinguistics and information theory, this work contributes to the ongoing debate on speaker-oriented and listener-oriented views of referential production. While previous research has primarily examined property selection and overspecification, a case that cannot strictly distinguish between these views, this thesis focuses on property ordering that renders syntactic variation. Because the listener-oriented hypothesis for property ordering requires greater production effort and influences the early stage of grammatical encoding during language production processing, it offers a clearer test case for differentiating the two views.

The findings from five experiments demonstrate that speakers employ both speaker- and listener-oriented approaches in linearization strategies. While speaker-oriented, heuristic speakers seek production efficiency by consistently relying on the preferred syntactic structure, listener-oriented speakers exhibit flexibility in ordering choices involving syntactic variation. The linearization pattern of this group is influenced not only by speaker-internal factors such as linguistic preferences and priming, but also, critically, by informativity. The preference for informative-first linearization provides robust evidence for listener-oriented communication efficiency, as placing the more informative property earlier significantly narrows the selection scope of potential referents for listeners. Another novel finding is that the proportion of listener-oriented speakers increases with communication engagement, suggesting that perspective-taking plays a mediating role in determining whether speakers adopt a speaker-oriented or listener-oriented production strategy.

Beyond advancing our understanding of referential production, this research offers methodological contributions. The maze-based sentence completion task helped to isolate the effect of RER on speakers' initial selection of referential properties, mitigating the potential occlusion of these effects in the less involving online experiments. The link between informativity and RER provides a quantifiable measure for interpreting the informativity of linguistic elements within visually-situated contexts. Additionally, these experiments carefully counterbalance potential confounding factors such as visual complexity and salience, which have not

always been rigorously controlled in prior research. The referential communication task also provides a structured prototype for studying information transmission in spontaneous, context-dependent, and task-constrained communication systems, contrasting with information-theoretic studies that focus on probabilistic trends of linguistic forms and structures.

Future research should further explore the interactive nature of communication by examining how both interlocutors dynamically adjust their strategies in real-time dialogue rather than focusing on a single role. Investigating more complex syntactic structures, such as constituent structures of objects, could help generalize the effect of informativity on linearization. Finally, different communication modalities, such as HCC versus HHC, may provide new insights into the cognitive aspects of communication,such as perspective-taking and alignment.

Appendices

A

Timing of Reference Production in Experiments 4 and 5

This appendix reports the experimental results concerning the timing of speakers' referential expression production. Previous research has examined the temporal dynamics of referential language through measures such as speech onset latency (e.g., Brown-Schmidt & Konopka, 2011; Gann & Barr, 2012; Vogels et al., 2020) and the duration between linguistic segments (e.g., Li et al., 2022). These measures serve as indicators of the cognitive effort and processing load associated with utterance planning. Longer latencies or durations typically reflect increased conceptual planning demands, more complex linguistic encoding, or articulation difficulty.

In the current project, the primary measure of interest was the first property selected by speakers, as this decision determined the syntactic structure of the entire utterance (e.g., selecting the animal property first resulted in a post-nominal structure). Timing data related to this selection were also recorded in Experiments 4 and 5, to provide supplementary insight into the cognitive effort involved in planning different syntactic structures and in the two speaker groups. These timing data allow for exploratory examination of whether listener-oriented strategies, which are assumed to be cognitively more demanding (e.g, Horton & Keysar, 1996; Keysar et al., 1998), can be reflected in longer planning times, thus helping to further distinguish between speaker-oriented and listener-oriented production patterns. This analysis is also relevant to previous literature regarding the mixed results of timing concerning overspecification (see Chapter 2.1.3.1).

In Experiment 4, timing was measured as the **Reaction Time (RT)** in Step 1 of the Maze Task, specifically, the interval between the maze task appearing and participants' selection of the first property to begin their utterance. In Experiment 5, which involved spontaneous spoken production, the relevant measure was the **First Property Latency (FPL)**: the time between the marking of the target referent and the acoustic onset of the first referential property in the utterance.

Both RT and FPL are linked to the planning effort required to select the initial property, a choice that also entails a syntactic decision for the utterance. However, it is important to acknowledge that these two measures may not reflect a strictly identical production process, due to the distinct nature of the tasks. The Maze Task involved explicit, visual decision-making, which may include additional cognitive effort in evaluating and rejecting non-target options. In contrast, FPL in spontaneous speech includes lexical retrieval and articulation processes, which were not required in the Maze Task due to its use of predefined options. Despite these differences, both measures can offer valuable insights into the planning processes underlying syntactic linearization.

Two exploratory hypotheses guided the analysis:

1. Structural difficulty: As Chapter 2.2.2.5 argues, post-nominal structures may be cognitively more demanding to produce than the pre-nominal structures. Therefore, both RT and FPL are expected to be longer for post-nominal modifications compared to pre-nominal ones, regardless of experimental conditions.
2. Speaker group differences: across the five experiments of the project, two types of speakers were consistently observed: *Group Consistent*, who adhered to a single syntactic structure, and *Group Varied*, who modulated their structural use based on informativity. The former are interpreted as relying on speaker-oriented heuristics, while the latter demonstrate a more listener-oriented strategy, possibly involving perspective-taking and audience design (e.g., Clark & Murphy, 1982; Ferreira, 2019). If these strategies differ in the required cognitive demands, *Group Varied* speakers should show longer RTs and FPLs than *Group Consistent* speakers.

The following sections report timing results for Experiments 4 and 5 separately. As the timing analyses are primarily exploratory, the focus is on identifying general trends rather than testing confirmatory hypotheses. Linear mixed-effects regression analyses are reported selectively, only where they serve to clarify key patterns in the data.

A.1 Experiment 4: Reaction Times of Step 1

Recall that in Experiment 4, participants took turns with a simulated Confederate to complete a maze-based sentence completion task for target description. Each Participant Turn began with the presentation of a display containing ten figures, without any indication of the target. After 2000 ms, a red square appeared to highlight the target figure, and after another 2000 ms, Step 1 of the Maze Task was shown below the display. At this stage, participants were instructed to proceed through the maze in two steps to complete their target description. Reaction times (RTs) were recorded at Step 1, capturing the latency between the presentation of the Maze Task and the participant's selection of the first property.

A.1.1 Data Recording and Analysis

RTs were recorded in each critical trial as the time interval (in milliseconds) between the appearance of Step 1, i.e., presenting "Wo ist ____" ["Where is ____"] along with four options, and the participant's click on one of these options to initiate the referential expression.

Following standard practices in reaction time analysis (e.g., Miller, 1991), RTs were excluded as outliers if they fell outside 2 Standard Deviations (SDs) from the mean. This criterion has been shown to minimize statistical bias while retaining most of the data (Berger & Kiefer, 2021). A total of 19 trials (1% of critical data) were excluded based on this threshold.

To assess the effects of syntactic structures on planning time, a linear mixed-effects model was fitted to the data. The dependent variable was log-transformed RTs to normalize the right-skewed distribution typical of RT data. The fixed effects included: **`Condition`** (Animal-Informative vs. Action-Informative), **`Structure`** (Pre-nominal vs. Post-nominal), **`Group`** (*Group Varied* vs. *Group Consistent*). All predictors were sum-to-zero coded. The best model included the three main effects without interaction. The random effects included random intercepts of subjects and items. The best model is selected following the model selection procedure described in Chapter 4.1.

For clarity and interpretability, data visualizations and descriptive summaries in the following sections are based on raw RTs, while all statistical analyses are based on log-transformed RTs.

A.1.2 Results

Figure A.1 depicts the distribution of raw RTs across the two speaker groups: *Group Consistent*, who exclusively produced pre-nominal modifications in Experiment 4 (28 participants), and *Group Varied*, who showed syntactic variation based on RER (51 participants). The data are further separated by condition (Animal-Informative vs. Action-Informative) and syntactic structure (Pre-nominal vs. Post-nominal).

Descriptively, the figure shows that mean RTs were generally longer for post-nominal modifications (i.e., when participants selected the animal property first in Step 1), with an approximate difference of 200 ms compared to pre-nominal modifications, consistent across both conditions. However, it is important to note that sample sizes were not balanced between the two structures: based on the main results in Chapter 5.1, participants overall showed a strong preference for pre-nominal modifications, as indicated by the higher density of semi-transparent dots for pre-nominal structures in the figure.

A second observation is that *Group Consistent* exhibited overall faster RTs than *Group Varied*, even when both groups produced pre-nominal structures in the same condition. Lastly, within *Group Varied*, RT differences of the two structures across the two conditions appeared relatively similar in both conditions.

To further evaluate these trends statistically, Table A.1 summarizes the results of the linear mixed-effects regression analysis. A significant main effect of `Group` was observed, indicating that RTs in *Group Varied* were significantly longer

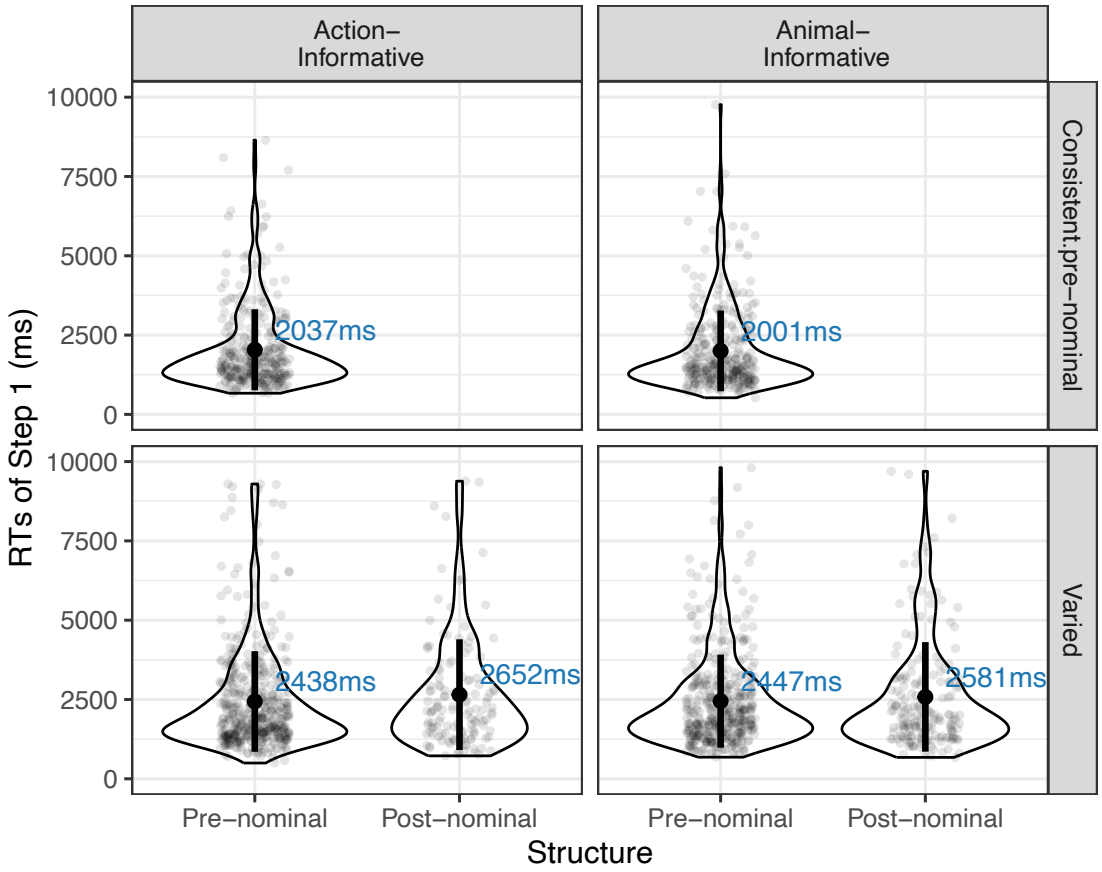


Figure A.1: Violin plots of Reaction Times (RTs) in Step 1 of Experiment 4, separated by condition (Animal-Informative vs. Action-Informative), syntactic structure (Pre-nominal vs. Post-nominal), and speaker group (*Group Varied* vs. *Group Consistent*). Note that *Group Consistent* in the figure included 28 subjects and *Group Varied* consisted of 51 subjects based on the main results of Experiment 4 (Chapter 5.1). Each violin includes semi-transparent black dots representing individual trials, with solid black points indicating the mean RT per condition per structure. Error bars represent the standard error of the mean. Raw RTs are shown in milliseconds.

than those in *Group Consistent*. No significant main effects of `Condition` or `Structure` were found.

A.1.3 Discussion

Two main observations emerge from the descriptive and statistical results of Experiment 4. First, the analysis revealed significantly longer RTs for *Group Varied* compared to *Group Consistent*, even when both groups produced the same pre-nominal structure under identical conditions. This finding suggests that adhering to a fixed syntactic structure, as seen in *Group Consistent*, potentially involved less cognitive effort during first property selection in Step 1 of the Maze Task. This supports the interpretation that repeatedly using the preferred pre-nominal structure reflects a heuristic strategy in referential production. In contrast, *Group*

Table A.1: Experiment 4. Statistical result of the main effects of Condition, Structure and Group on log transformed RTs

Predictor	$\hat{\beta}$	95% CI	t	df	p
Intercept	7.55	[7.48, 7.63]	192.19	101.48	$< .001$
Structure	0.02	[-0.02, 0.05]	1.03	1809.20	.301
Condition	0.00	[-0.02, 0.02]	0.14	1750.18	.887
Group	-0.11	[-0.17, -0.04]	-2.98	78.90	.004

Note. The model was specified as: RT ~ Condition + Structure + Group + (1|subject) + (1|item). All predictors were sum-to-zero coded. Each coefficient represents a deviation from the grand mean.

Varied, despite producing the same pre-nominal structure, appeared to engage more effort in determining the first property and so the corresponding syntactic structure, reflected in their longer RTs.

Second, post-nominal modifications were associated with longer RTs than pre-nominal ones, although this difference did not reach statistical significance. One possible explanation is the unequal sample sizes across structures, resulting from participants' general preference for pre-nominal modifications. Nonetheless, this descriptive trend provides additional support for the idea that producing post-nominal structures may require greater planning and production effort, as discussed in Chapter 2.2.2.5.

The next section reports the First Property Latency (FPL) results from Experiment 5, which were expected to replicate the RT patterns observed in Experiment 4 in general.

A.2 Experiment 5: First Property Latency

In Experiment 5, participants produced spoken target descriptions to guide a Confederate in selecting the target figure. The utterances for each trial were recorded individually, starting from the moment when the red square appeared to highlight the target. First Property Latency (FPL) was defined as the time interval between the marking of the target and the onset of the first referential property in the utterance.

Rather than measuring speech onset, i.e., the point at which the participant began speaking, typically just before uttering "clicking on...", FPL provides a more precise estimate of the planning effort involved in selecting the first content word. This choice was based on the assumption that speakers may initiate speech before fully planning the content of their description, given that language production is an incremental process (see Chapter 2.2.2.1). For example, they may begin with the frame "clicking on the..." while still deciding how to describe the target referent,

during which disfluencies such as fillers (e.g., "uh"), pauses, and hesitations may also occur. If only speech onset were measured, the timing reflecting this ongoing planning may be missed. FPL includes this extended timing, capturing the point at which speakers settle on the first property and thereby determine the syntactic structure of the referential expression. It offers a broader estimate of the planning effort involved in utterance formulation.

A.2.1 Data Recording and Analysis

Participants' utterances were recorded as separate audio files per trial and subsequently transcribed. Speech disfluency, including long pauses, fillers (e.g., "uh"), hesitations, and repairs, was annotated during transcription.

To extract the timing data, the Montreal Forced Aligner (MFA) (McAuliffe et al., 2017) was used. MFA is an open-source tool that time-aligns transcriptions with their corresponding audio files using pretrained acoustic models and pronunciation dictionaries. The raw FPL was defined as the timestamp of the first property word as returned by MFA.

Following disfluency coding, a trial was excluded under the following scenarios:

1. A cough or non-task-related vocalization by the participants or the Confederate occurred before the first property.
2. A lexical repair occurred before the first property was produced (e.g., "den Hund, der weint – uh den den Löwen, der weint" [the dog that cries – uh the lion that cries]), or a sentence-level restart occurred. However, stuttering (e.g., "den ma–malenden Hasen" [the painting rabbit]) was retained if it indicated that the speaker had already selected the first property. In such cases, the FPL was measured from the onset of the first stutter (i.e., before "ma–").
3. The participant omitted the sentence frame (e.g., "klick auf _____" [click on _____]) and produced only a nominal phrase, which resulted in substantially earlier onsets that were not comparable to other trials.
4. Other technical issues or interruptions occurred at trial onset.

A total of 47 trials (2.1% of critical trials) were removed at this stage based on the above criteria.

To further remove outliers, FPLs falling outside 2 SDs from the mean were excluded. This resulted in the exclusion of an additional 99 trials (4.6%) for subsequent analysis.

Similar to Experiment 4, a linear mixed-effects model was conducted. The dependent variable was log-transformed RTs to be consistent with Experiment 4. The fixed effects included: **`Condition`** (Animal-Informative vs. Action-Informative), **`Structure`** (Pre-nominal vs. Post-nominal), **`Group`** (*Group Varied* vs. *Group Consistent*). All predictors were sum-to-zero coded. The best model included the three main effects without interaction. The random effects included random intercepts and random slopes of `Condition` for subjects, as well as random intercepts of items. The best model is selected following the model selection procedure described in Chapter 4.1.

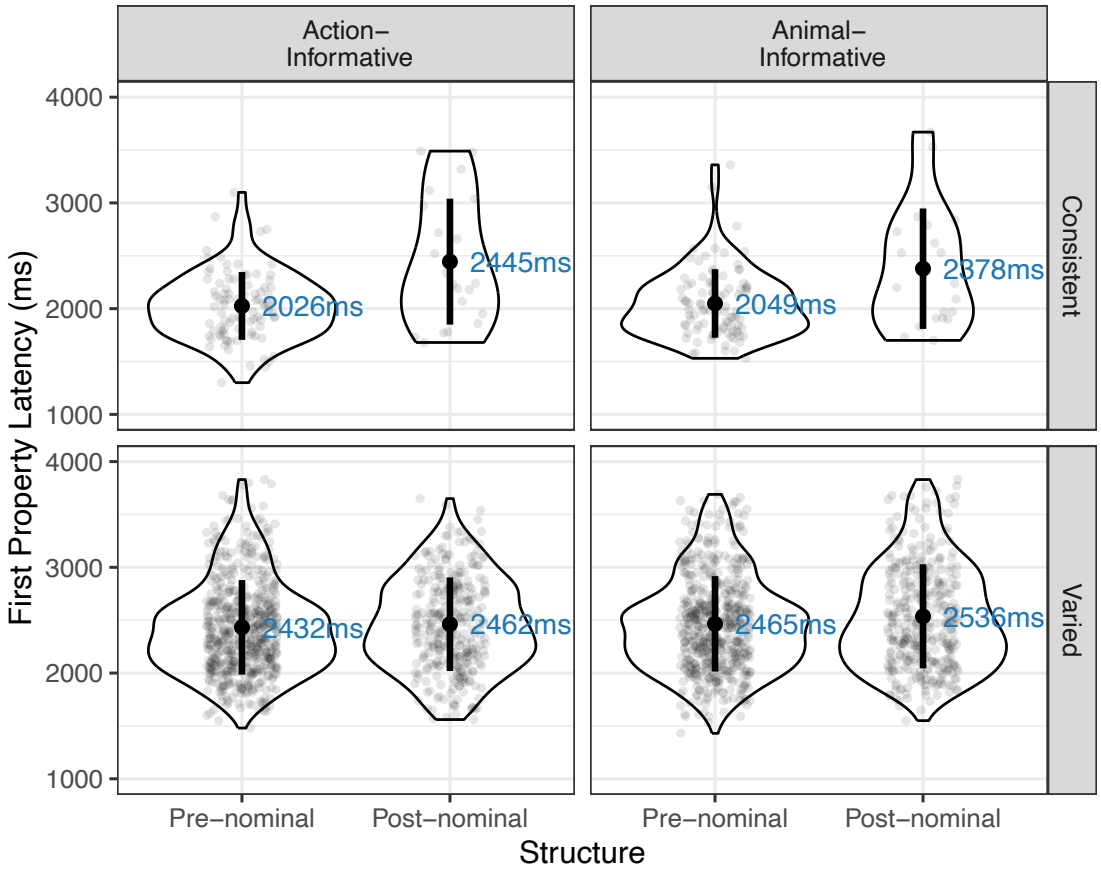


Figure A.2: Experiment 5. Violin plots of First Property Latency (FPL) in Experiment 5, separated by condition (Animal-Informative vs. Action-Informative), syntactic structure (Pre-nominal vs. Post-nominal), and speaker group (*Group Varied* vs. *Group Consistent*). Note that *Group Consistent* included four participants who only produced pre-nominal modifications and one participant who consistently produced post-nominal modifications, while *Group Varied* comprised 40 participants. Each violin includes semi-transparent black dots representing individual trials, with solid black points indicating the mean latency per condition. Error bars represent the standard error of the mean. Raw FPL values are shown in milliseconds.

A.2.2 Results

Figure A.2 illustrates the distribution of raw FPL across speaker groups, conditions, and syntactic structures. The descriptive results were similar to Experiment 4: *Group Varied* speakers showed longer FPLs overall than *Group Consistent*, and post-nominal modifications were associated with longer latencies than pre-nominal ones.

To further evaluate these trends statistically, Table A.2 summarizes the results of the linear mixed-effects regression analysis. A significant main effect of `Group` was observed, indicating that FPLs in *Group Varied* were significantly longer than those in *Group Consistent*. A main effect of `Structure` was also detected, suggesting that producing post-nominal modifications resulted in significantly longer

Table A.2: Experiment 5. Statistical result of the main effects of Condition, Structure and Group on log transformed FPLs

Predictor	$\hat{\beta}$	95% CI	t	df	p
Intercept	7.72	[7.67, 7.78]	289.59	46.87	$< .001$
Structure	-0.02	[-0.03, -0.01]	-3.85	1989.19	$< .001$
Condition	-0.01	[-0.01, 0.00]	-2.04	42.91	.048
Group	-0.08	[-0.13, -0.03]	-2.96	43.12	.005

Note. The model was specified as: FPL ~ Condition + Structure + Group + (1+Condition|subject) + (1|item). All predictors were sum-to-zero coded. Each fixed-effect coefficient represents a deviation from the grand mean.

FPLs than producing pre-nominal ones.[1] Lastly, a main effect of `Condition` was observed: producing referential expressions in the Animal-Informative Condition also led to longer FPLs than in the Action-Informative Condition. [2]

A.2.3 Discussion

The results of Experiment 5 revealed a pattern consistent with Experiment 4 regarding the timing of referential expression production. First, FPLs for *Group Consistent* speakers were significantly shorter than those for *Group Varied* speakers, especially for pre-nominal modifications. This difference supports the interpretation that repeatedly using the preferred pre-nominal structure reflects a heuristic strategy in referential production. The contrast is especially clear when comparing FPLs for pre-nominal modifications: although both groups produced the same structure, *Group Varied* speakers exhibited significantly longer latencies. This suggests that they engaged in more cognitively effortful planning, possibly because they adopted a listener-oriented strategy that considered the informativity of properties and the listener's needs during utterance formulation. In contrast, *Group Consistent* speakers followed a speaker-oriented, heuristic approach that required less planning effort.

[1]This effect was not driven by the one participant in *Group Consistent* who consistently produced post-nominal structures throughout the experiment. A separate analysis excluding all five participants from *Group Consistent* (i.e., restricting the analysis to *Group Varied*) yielded the same main effects of `Structure` and `Condition`. As this analysis revealed the same overall trends, it is not reported in detail.

[2]This effect may be partially driven by the higher number of post-nominal productions in the Animal-Informative Condition, which is an expected outcome based on one of the central predictions of the current project, namely the informative-first linearization preference. However, this distributional pattern would inevitably introduce collinearity between the two predictors `Condition` and `Structure`, meaning that both potentially account for overlapping variance. Therefore, the interpretation of these main effects should be treated with caution.

Second, post-nominal modifications were associated with significantly longer FPLs in Experiment 5. Unlike in Experiment 4, this difference reached statistical significance, probably due to a larger number of trials registered with post-nominal structures. In the spoken production task, post-nominal structures accounted for approximately 37% of utterances, compared to only about 20% in the maze-based task (see Chapter 5). It also provided supporting evidence that producing post-nominal structures may require greater planning effort than pre-nominal structures, consistent with the discussion in Chapter 2.2.2.5

A.3 General Discussion and Conclusion

This appendix reports the timing dimensions of referential production in Experiment 4, using a Maze Task, and in Experiment 5, using a spoken production task. In Experiment 4, timing was measured as the RTs taken by participants to select the first property in Step 1 of the Maze. In Experiment 5, it was measured as the First Property Latency (FPL), the duration between the presentation of the target and the onset of the first referential property in speech. Although these two measures do not reflect strictly identical production processes due to inherent differences between the tasks, both can be interpreted as indicators of planning effort during the selection of the first property, which further determines the syntactic structure of the entire utterance. The exploratory results across both tasks converge on two central claims developed in the main body of this thesis.

First, both RTs and FPLs showed that *Group Consistent* speakers exhibited significantly shorter planning times than *Group Varied* speakers when producing referential expressions. As discussed in the main chapters (Chapter 2.1.1), *Group Consistent* speakers were associated with a speaker-oriented view of referential production: they did not systematically consider informativity or the listener's needs, but instead relied on what was the easiest for them for production, by repeatedly using the preferred pre-nominal structure. The shorter planning times observed for this group offer direct empirical support for this interpretation: repeatedly using the preferred syntactic structure, regardless of informativity, is indeed a heuristic strategy that minimizes cognitive effort.

In contrast, *Group Varied* speakers, who were linked to a listener-oriented approach, modulated their structure choices based on informativity and the needs of the listener. This group expended longer timing, thereby potentially more cognitive effort, during utterance planning, when selecting the first property. Notably, *Group Varied* speakers showed significantly longer RTs and FPLs even when producing the same pre-nominal structure as *Group Consistent* speakers, by over 400 ms in both experiments. This strongly suggests that their increased planning time was not due to structural complexity per se, but rather to the additional cognitive load involved in audience design. These findings also align with previous claims that listener-oriented processing imposes greater demands on the speaker (e.g., Keysar et al., 1998).

Second, the timing results from both experiments provide converging evidence that post-nominal modifications require more planning effort than pre-nominal ones.

This supports the claim, discussed in Chapter 2.2.2.5, that producing post-nominal structures involves a qualitatively different and more difficult planning process. Importantly, despite this increased difficulty, *Group Varied* speakers continued to modulate their structure choices based on informativity. This finding reinforces the interpretation that listener-oriented speakers invest additional effort to support efficient and incremental target identification for the listener, even when this entails producing the less preferred and more complex syntactic form.

B

Power Analysis of Experiment 1 and Experiment 2, Using `Simr` R Package

This appendix presents a simulation-based power analysis for Experiment 1 and Experiment 2 using the R package `simr` (Green & MacLeod, 2016). The analysis had two primary goals.

First, the power analysis aimed to determine how many participants and how many critical trials per participant are required to reliably detect the effect of informativity across the Animal-Informative and Action-Informative Conditions. These two conditions were used consistently throughout the project, and identifying the optimal balance between participant count and trial number was crucial for informing the design of later experiments, particularly Experiments 4 and 5. Comparing the power of Experiment 1 and Experiment 2 provides an ideal contrast between trial-rich (Experiment 1) and participant-rich (Experiment 2) designs, offering insights into how these design choices affect statistical sensitivity to the effect of informativity.

Second, the analysis aimed to clarify an interpretive issue concerning the statistical results reported in the project. While both Experiments 1 and 2 yielded robust informativity effects within the *Group Varied* participants, only Experiment 2 showed a significant effect in the overall sample. Experiment 1 employed a trial-rich design, with 79 participants (35 in *Group Varied*), each completing 24 critical trials. In contrast, Experiment 2 used a participant-rich design, recruiting 149 participants (70 in *Group Varied*), each completing only 12 critical trials.

This discrepancy in both design and results raised the question of whether the observed significance in Experiment 2 truly reflects an improvement in experimental design (i.e., due to adding the Listener Task block), or whether it is primarily a consequence of the larger sample size, which inherently increases the likelihood of obtaining a statistically significant effect (e.g., Andrade, 2020). Power simulations

were used to address this concern by evaluating which experimental structure (trial-rich or participant-rich) would reliably detect a theoretically meaningful effect when extended to a large enough number of participants or trials.

To answer these questions, simulation-based power estimation was conducted using the simr package (Green & MacLeod, 2016), an R package designed for power analysis in regression models with random effects—that is, linear mixed models (LMMs) or generalized linear mixed models (GLMMs), such as the logistic regression models used in the current analysis. These models were fitted using the `lmer()` and `glmer()` functions from the `lme4` package (Bates et al., 2015). The analysis involves generating simulated datasets based on an observed fitted model (i.e., models based on Experiments 1 and 2) and systematically varying design parameters such as the number of participants or items, and estimating how often the effect of interest reaches statistical significance across multiple simulated replications.

Statistical power refers to the probability that a test will correctly reject a false null hypothesis. Ensuring sufficient power—typically set at 80%—is critical, as it reflects the design's ability to detect real effects when they exist. In `simr`, power is more specifically calculated as the proportion of simulated datasets in which the fixed effect of interest (here, `Condition`) reaches statistical significance, given a specified effect size and model structure. For example, a power estimate of 80% means that the effect was significant ($p < .05$) in approximately 800 out of 1,000 simulations. This simulation-based method provides a direct estimate of how reliably a given design can detect a true effect that is theoretically meaningful.

B.1 Method

According to Green & MacLeod (2016), the simulation-based power analysis procedure involves three general steps:

1. **Fitting a model from observed data**: A (generalized) LMM was fitted to the observed data from each experiment. This model included fixed effects (e.g., `Condition`) and random intercepts for participants and items, reflecting the empirical structure of the experiments.

2. **Assigning an estimated effect size**: The fixed effect of interest (here, the effect of `Condition`) should be estimated by the experimenters and be assigned a theoretically meaningful value before simulation. A common and conservative approach for this estimation is to find the smallest effect size of interest (SESOI), the minimum effect size that is significant or theoretically worth detecting (Albers & Lakens, 2018; Kumle et al., 2021). If the goal of the analysis is to estimate the power of the current model without further modifying it, a basic power analysis can already be conducted directly on the current fitted model using `simr`.

3. **Increasing the number of participants and/or items**: The model with the estimated effect size was then used to simulate new datasets under varying sample sizes, either by extending the number of participants and/or the number of critical trials. For each simulated dataset, the model was re-fitted, and the fixed

effect was tested for statistical significance. Power was estimated as the proportion of simulations in which the effect reached significance at $p < .05$.

In the current analysis, which contrasts trial-rich (Experiment 1) and participant-rich (Experiment 2) designs, the number of participants was extended in Experiment 1, while the number of trials per participant was extended in Experiment 2. These simulations address complementary questions about design parameters of the number of participants and items:

- For Experiment 1, given 24 critical items, how many participants are required to achieve sufficient power?
- For Experiment 2, given 70 participants, how many items are needed per participant to reach the same level of power?

Power curves are generated to visualize how power changed as a function of sample size, to estimate the required number of participants or items for future research.

B.2 Analysis

To begin the power analysis from Step 1, logistic mixed-effects regression models were fitted to the data from *Group Varied* participants in Experiments 1 and 2 (see Chapter 2.1.2.2). In both models, the dependent variable was `Structure` (pre-nominal vs. post-nominal), and the predictor was `Condition` (Animal-Informative vs. Action-Informative).

For simplicity and model stability, the random-effects structure in both models included only random intercepts for subjects and items. The predictor `Condition` was dummy-coded, with the Action-Informative Condition as the reference level. This coding allowed the fixed effect of `Condition` to be interpreted directly as the increase in post-nominal structure use in the Animal-Informative Condition relative to the Action-Informative baseline.

```
Structure ~ Condition +(1|PP) +(1|display)
```

For Step 2, the same estimated effect size was assigned to both models. For both models, the fixed effect of `Condition` was manually set to $\beta = 0.4$ on the log-odds (logit) scale, following the smallest effect size of interest (SESOI) approach. This value reflects a modest but theoretically meaningful effect size: it is large enough to yield statistical significance, yet conservative relative to the observed effects in the current project ($\beta \approx 1.08$ in Experiment 1 and $\beta \approx 0.47$ in Experiment 2).

On the probability scale, a logit difference of 0.4 corresponds to approximately a 7% increase in the likelihood of producing post-nominal modifications between the two conditions. This magnitude is consistent with the range of effects observed across the five experiments reported in this dissertation.

For both models, $\beta_{intercept}$ was set to -1.39, which corresponds to an estimation of 20% post-nominal use in the Action-Informative Condition as the reference level, based on the mean proportion of post-nominal modifications in the Action-Informative Condition across the two experiments.

For Step 3, the sample sizes of participants and items were increased using the `extend()` function from the `simr` package. For Experiment 1 (which originally included 24 items), the fitted model was extended to include up to 100 participants (from an initial 35 *Group Varied* participants), in order to identify the number of participants required to achieve 80% power.

For Experiment 2, the number of trials per participant was extended from 12 to 72, while keeping the number of *Group Varied* participants fixed at 70, to estimate how many items were needed to reach sufficient power.

Power curves were generated using the `powerCurve()` function, each based on 1,000 simulation runs, to illustrate how power changes as a function of increasing participants (Experiment 1) or trials (Experiment 2), which will be reported in the results below.

B.3 Results

Figure B.1 illustrates how power increases as a function of the number of *Group Varied* participants for the model based on Experiment 1. Power reaches the conventional 80% threshold with approximately 50 participants, assuming 24 critical trials per participant.

Figure B.2 shows how power changes as a function of the number of critical trials per participant for the model based on Experiment 2. Power reaches 80% with approximately 36 trials per participant, assuming a fixed sample of 70 *Group Varied* participants.

Across the two simulations, the models reached 80% power under different design adjustments. Experiment 1 achieved this threshold with approximately 50 participants, while Experiment 2 required 36 trials per participant. These values offer a descriptive comparison of trial-rich versus participant-rich designs under a consistent effect size assumption. Notably, Experiment 2 required both a larger sample size and more critical trials per participant to reach the same level of power as Experiment 1.

B.4 Discussion

By conducting two power analyses based on the observed data of Experiments 1 and 2 using the `simr` package, this appendix identified how many participants and trials are needed to achieve sufficient power for detecting informativity effects using the conditions and tasks of the current project.

The results showed that Experiment 1 achieved greater design efficiency per participant due to its higher trial count, while Experiment 2 required substantially more items to reach comparable power levels. This highlights the importance of within-subject repeated measures for detecting the effects of informativity with a relatively modest effect size.

It is not surprising that Experiment 2 required both more participants and more trials to reach 80% power compared to Experiment 1. This difference can be

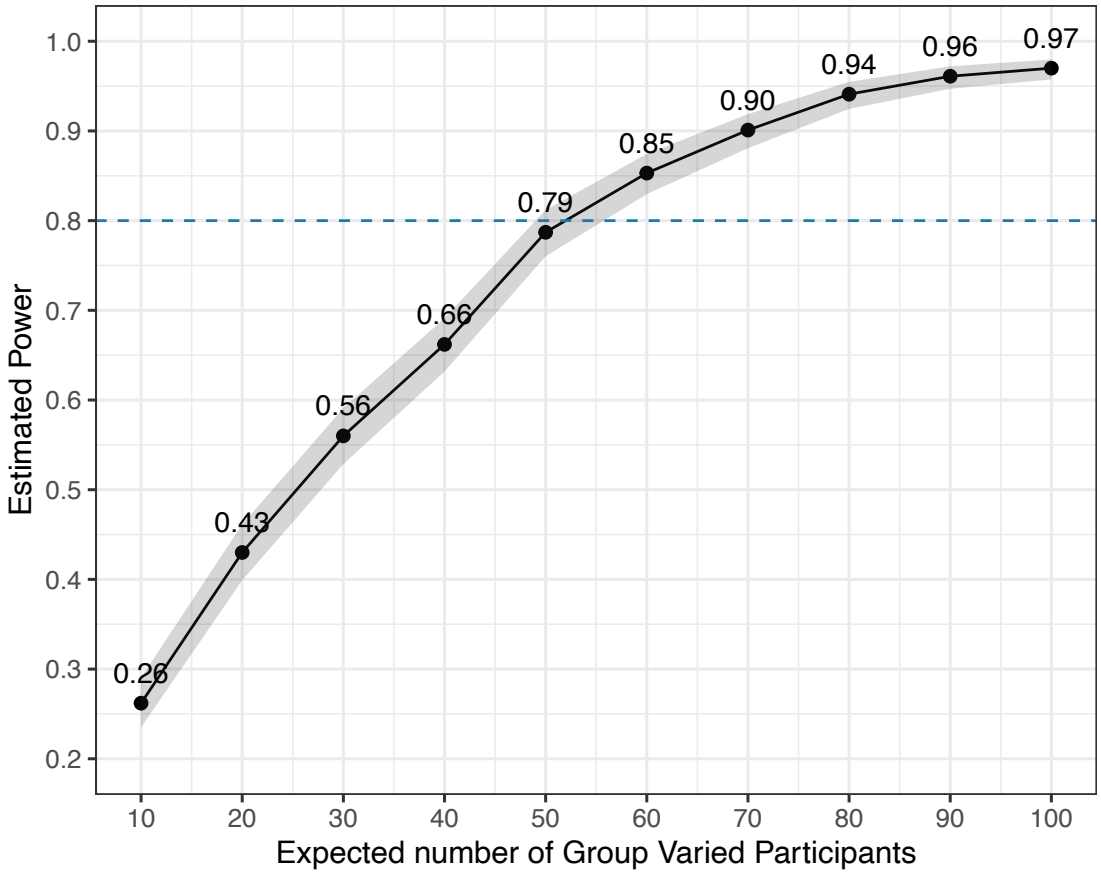


Figure B.1: Estimated power as a function of the number of *Group Varied* participants in Experiment 1, assuming 24 critical trials per participant. Power was estimated using the `simr` package with 1,000 simulation runs at each sample size. The grey ribbon indicates the 95% confidence interval of the simulated power estimates, and the blue dashed line marks the conventional 80% power threshold.

partly attributed to the design: each participant in Experiment 2 completed only 12 critical items, drawn from two versions of the experimental lists (each containing 12 different items), whereas each participant in Experiment 1 completed the full set of 24 critical items. Additionally, the effect size detected in Experiment 2 was smaller ($\beta \approx 0.47$) than in Experiment 1 ($\beta \approx 1.1$), further reducing its statistical sensitivity. While both power simulations used a fixed effect size of 0.4 for comparability, the empirical structure of Experiment 2 may introduce greater between-subject variability and less within-subject repetition that may reduce power. This is also partially reflected in the model of Experiment 2 reported in Chapter 4.2, which included random slopes of both participants and items to account for this variability (however, to ensure a stable and interpretable comparison, models adopted for the current analysis were based on simplified models with random intercepts only). All of these concerns reinforce the conclusion that the trial-rich design of Experiment 1 yielded greater statistical power and design efficiency for detecting the effect of informativity.

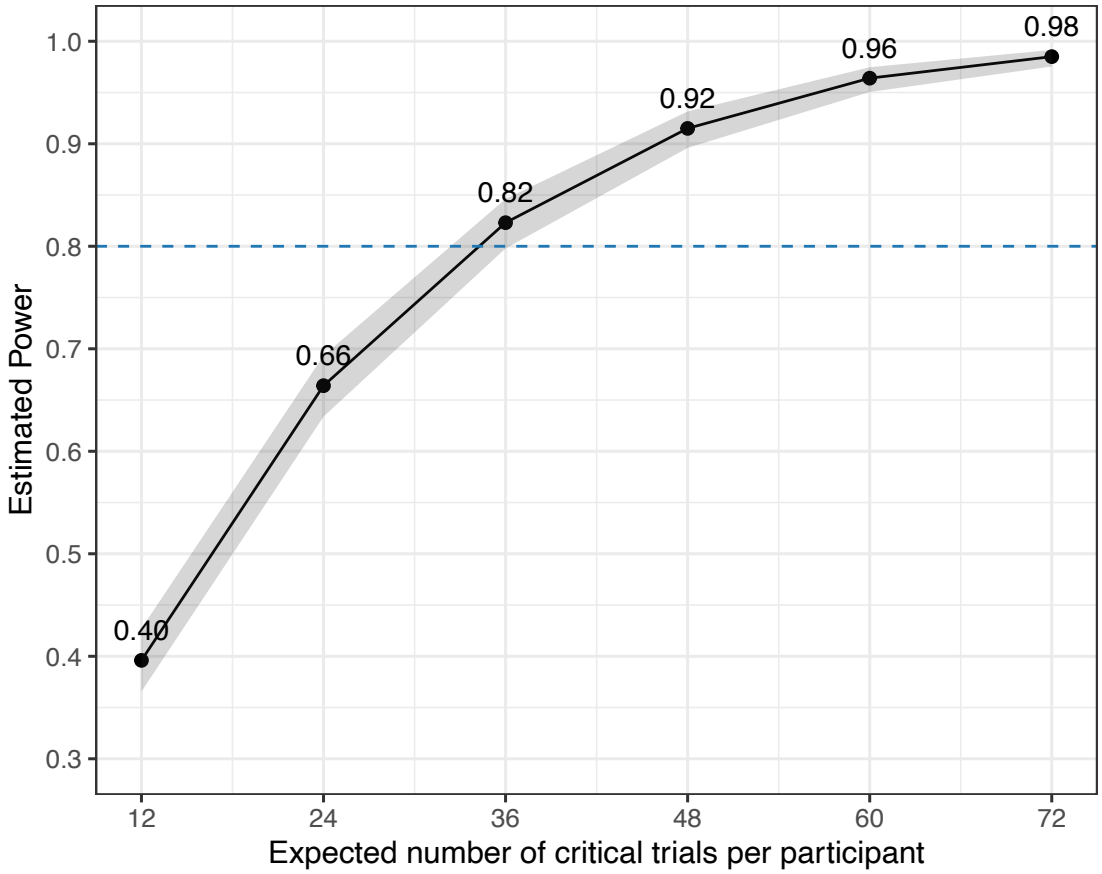


Figure B.2: Estimated power as a function of the number of critical trials per participant in Experiment 2, assuming 70 *Group Varied* participants. Power was estimated using the `simr` package with 1,000 simulation runs per trial count. The grey ribbon indicates the 95% confidence interval of the simulated power estimates, and the blue dashed line marks the conventional 80% power threshold.

It is also important to note that these power estimates do not imply that the significant effects reported in Chapter 2.1.2.2 were invalid or underpowered, especially in Experiment 1. The effect size detected in Experiment 1 was substantially larger than the simulated SESOI value of 0.4, making the result robustly significant even with a smaller sample.

These findings also address the second goal of this appendix: to evaluate whether the significant effect observed in the overall model of Experiment 2 reflects a genuinely more sensitive experimental design or is simply a consequence of increased sample size. The results suggest the latter.

In Experiment 1, only the *Group Varied* subset (35 participants and 24 trials) produced a significant effect, but the overall model (79 participants) did not. This is probably because *Group Consistent* participants, who did not vary their structure use across conditions, diluted the effect in the overall model.

However, Experiment 2 yielded a statistically significant effect in both the *Group Varied* model (70 participants and 12 trials) and the overall model, including all 149 participants. According to the power curve in Figure B.2, the fixed effect of

Condition in Experiment 2 would require approximately 36 trials per participant to achieve 80% power with 70 *Group Varied* participants. The actual design, which included only 12 critical trials per participant, thus falls short of this threshold. The additional 79 *Group Consistent* participants did not vary their structural choices across conditions and therefore did not contribute meaningful variation to the effect of interest. However, their inclusion increased the total number of observations and reduced standard errors in the overall model, thereby increasing the likelihood of obtaining a statistically significant result (e.g., Andrade, 2020).

Therefore, the significant result in the overall model in Experiment 2 (as well as for Experiment 3 following an identical design) should be interpreted with caution: it is more likely to be a statistical artifact of increased sample size rather than genuine evidence of a stronger experimental manipulation.

In sum, the power analysis presented in this appendix demonstrated how the total number of participants and trials interacts with statistical power in the context of the current project. Using the `simr` package, the analysis highlighted the use of trial-rich designs for achieving more reliable and adequately powered experiments, particularly when the effect size is expected to be small to moderate.

C

Zusammenfassung in deutscher Sprache

Diese Dissertation untersucht den Einfluss von Informativität auf die syntaktische Linearisierung von Objekteigenschaftenin der referenziellen Produktion. In visuell eingebetteten referenziellen Kommunikationst-Tasks konstruieren Sprecher Ausdrücke, um einem anwesenden Hörer zu helfen, ein spezifisches Zielobjekt in einer visuellen Szene zu identifizieren. Dieser Prozess umfasst zwei zentrale Komponenten: 1) die Eigenschaftsauswahl, also die Auswahl relevanter Eigenschaften, die eine eindeutige Zielidentifikation unterstützen, und 2) die Eigenschaftsreihenfolge, also die Anordnung der gewählten Eigenschaften in einer syntaktisch linearisierten Form. Während frühere Studien besonderes Augenmerk auf Überspezifizierung (ÜS) bei der Eigenschaftsauswahl gelegt haben – bei der Sprecher mehr Informationen einfügen als notwendig (z.B. ein blaues Dreieck", obwohl nur ein Dreieck vorhanden ist) – ist wenig darüber bekannt, wie Eigenschaftsreihenfolge als Linearierungsprozess in der referenziellen Produktion funktioniert, insbesondere wenn syntaktische Variation möglich ist, etwa bei prä-nominalen versus post-nominalen Strukturen (z.B. *der weinende Hase* vs. *der Hase, der weint*).

Indem referenzielle Kommunikation als kooperativer Prozess (Clark & Wilkes-Gibbs, 1986; Grice, 1975) und als ein Akt der Informationsübertragung (Sperber & Wilson, 1995) verstanden wird, der auf Prinzipien der Informationstheorie beruht (Shannon, 1948), untersucht diese Dissertation, ob die referenziellen Äusserungen von Sprechern durch Informativität im Sinne kommunikativer Effizienz beeinflusst werden. Genauer gesagt wird erforscht, ob Informativität die Linearisierung von referenziellen Eigenschaften beeinflusst, wenn diese Linearisierung mit syntaktischen Variationen in der referenziellen Enkodierung einhergeht. Das Ziel, den Einfluss von Informativität auf syntaktische Linearisierung zu untersuchen, ist es, zur theoretischen Debatte zwischen sprecherorientierten und hörerorientierten Sichtweisen der referenziellen Produktion beizutragen, indem ein Untersuchungsgegenstand behandelt wird, der über bisherige Forschung zur ÜS hinausgeht.

Das Projekt formuliert drei Forschungsfragen.

Forschungsfragen 1: Beeinflusst Informativität die syntaktische Linearisierung in der referenziellen Produktion?

Forschungsfragen 2: Falls Informativität eine Rolle spielt, wie moduliert sie die Reihenfolge referenzieller Eigenschaften?

Forschungsfragen 3: Falls sich ein konsistentes Linearisierungsmuster zeigt, das durch Informativität beeinflusst wird, verstärkt kommunikative Einbindung diesen Effekt?

Diese Fragen werden in fünf Experimenten untersucht, die die Informativität referenzieller Adjektive und Substantive in visuell eingebetteten referenziellen Kommunikationstasks manipulieren.

Das Projekt operationalisiert Informativität mithilfe der Referential Entropy Reduction (RER, Tourtouri et al., 2019), die quantifiziert, wie stark ein Eigenschaftswort die Unsicherheit bei der Identifikation eines Zielreferenten im visuellen Kontext reduziert. RER ist umgekehrt proportional zum Auswahlbereich – der Anzahl potenzieller Referenten, auf die ein Eigenschaftswort zutrifft. Ein informatives Wort mit höherer RER reduziert die Unsicherheit stärker, da es sich auf eine kleinere Menge von Referentenkandidaten in der Szene bezieht.

Kapitel 2 dieser Arbeit fasst frühere Forschung zur referenziellen Produktion zusammen und zeigt auf, dass der bisherige Schwerpunkt auf Eigenschaftsauswahl – insbesondere auf ÜS – um eine genauere Untersuchung der Informativität und der Eigenschaftsreihenfolge erweitert werden sollte. ÜS ist ein zentrales Thema in der Forschung zur referenziellen Kommunikation, bei dem Sprecher häufig mehr Informationen angeben, als für die Identifikation des Ziels notwendig wären. Sie wurde verwendet, um sowohl sprecherorientierte als auch hörerorientierte Sichtweisen der referenziellen Produktion zu stützen.

Die sprecherorientierte Sichtweise (Kapitel 2.1.1) argumentiert, dass ÜS häufig auftritt, weil referenzielle Ausdrücke heuristisch und egozentrisch produziert werden (z.B. Horton & Keysar, 1996; Keysar et al., 1998): Sprecher priorisieren sprecherinterne Faktoren wie Produktionsökonomie (MacDonald, 2013) und berücksichtigen selten die Bedürfnisse des Hörers. Sie konzentrieren sich auf die Eigenschaften des Zielobjekts selbst, ohne es mit Distraktoren zu vergleichen – insbesondere in komplexen visuellen Kontexten (z.B. Koolen et al., 2013) – und bevorzugen visuell saliente Eigenschaften sowie syntaktische Strukturen, die häufig und leicht zugänglich sind (z.B. Gatt et al., 2013; Tarenskeen et al., 2015).

Im Gegensatz dazu geht die hörerorientierte Sichtweise (Kapitel 2.1.2) davon aus, dass der Sprecher die Perspektive des Hörers während der referenziellen Produktion berücksichtigt, z.B. im Sinne von Adressatenorientierung (Clark & Murphy, 1982). Sprecher wissen, dass der visuelle Suchprozess des Hörers inkrementell verläuft (Rubio-Fernández, 2016): Der Hörer interpretiert jedes Eigenschaftswort beim Eintreffen, wodurch sich die Menge möglicher Referenten zunehmend einschränkt (z.B. Eberhard et al., 1995; Rubio-Fernández et al., 2021). ÜS kann dabei helfen, das Ziel effizienter zu identifizieren, insbesondere wenn informative Eigenschaften früh auftreten. So fanden Rubio-Fernández (2016) beispielsweise, dass informative Farbüberspezifikstion im Englischen häufiger auftritt als im Spanischen, da die platzierte Eigenschaft im Englischen prä-nominal (*the blue dress* [das *blaue* Kleid]) effektiver für die Zielidentifikation ist als im post-nominalen Spanisch (*el vestido*

azul [das Kleid *blau*]). Tourtouri et al. (2019) zeigte, dass einige Sprecher die informativere Eigenschaft gemäss RER überpräzisierten, um die Zielidentifikation effizienter zu gestalten.

Obwohl beide Sichtweisen ÜS als zentrales Testfeld nutzen, unterscheiden sie sich darin, wie sie empirische Ergebnisse zum ÜS-Verhalten von Sprechern interpretieren. Ähnliche Muster – wie erhöhte Farbüberspezifizierung in komplexen Szenen – wurden entweder auf visuelle Salienz (sprecherorientiert, z.B. Koolen et al., 2013) oder auf Erleichterung der visuellen Suche (hörerorientiert, z.B. Rubio-Fernández, 2019) zurückgeführt. Diese kontrastierenden Interpretationen desselben empirischen Trends werfen die Frage auf, ob ÜS wirklich ein ideales Testfeld zur Unterscheidung der beiden divergierenden Sichtweisen darstellt.

Um dieses Problem zu lösen, verlagert das vorliegende Projekt den Fokus von ÜS auf die Eigenschaftsreihenfolge und die Rolle der Informativität bei der Bestimmung der Linearisierung. Wie in Kapitel 2.2 beschrieben, unterliegt die Eigenschaftsreihenfolge häufig grammatischen und semantischen Beschränkungen. Die Produktion weniger bevorzugter Reihenfolgen – wie post-nominaler Modifikationen – erfordert typischerweise höheren Produktionsaufwand (z.B. Brown-Schmidt & Tanenhaus, 2006) und könnte daher gezielte kommunikative Ziele widerspiegeln, etwa die Ausrichtung an einem Interaktionspartner (z.B. Cleland & Pickering, 2003) oder die Priorisierung von Informativität (z.B. Fukumura, 2018).

Diese Dissertation betrachtet Informativität als einen sprecherexternen Faktor, der durch kommunikative Effizienz motiviert ist und eng mit dem inkrementellen visuellen Suchprozess des Hörers verknüpft ist. In der breiteren Literatur zur Sprachproduktion gelten sprecherexterne Faktoren wie Adressatenorientierung und Perspektivübernahme häufig als weniger einflussreich als sprecherinterne Faktoren wie Verfügbarkeit oder syntaktische Präferenzen, wenn es darum geht, welche lexikalischen und syntaktischen Entscheidungen Sprecher treffen (z.B. Damen et al., 2019; Ferreira et al., 2005; Ferreira & Dell, 2000; Wardlow Lane et al., 2006).

Dies macht die Untersuchung des Einflusses von Informativität auf die Eigenschaftsreihenfolge zu einem starken Testfall. Wenn sich zeigt, dass Informativität die Eigenschaftsreihenfolge moduliert, stützt dies die hörerorientierte Sichtweise der referenziellen Produktion als kooperativer, effizienzgetriebener Prozess, bei dem Sprecher ihre Äusserungen auf die visuelle Suche des Hörers nach dem intendierten Ziel ausrichten. Umgekehrt würde ein konsistentes Festhalten der Sprecher an der bevorzugten syntaktischen Struktur – unabhängig von Informativität oder Kommunikationskontext – die sprecherorientierte Sichtweise stützen und nahelegen, dass die Phase der grammatischen Enkodierung in der Sprachproduktion unempfänglich für externe kommunikative Anforderungen bleibt.

Ausgehend von den Forschungsfragen und theoretischen Perspektiven formuliert das Projekt drei Hypothesen und entsprechende Alternativhypothesen für die Forschungsfragen 1 bis 3. Bezüglich Forschungsfrage 1 – ob Informativität grundsätzlich die syntaktische Linearisierung beeinflusst – sagt die hörerorientierte Sichtweise voraus, dass Sprecher die Eigenschaftsreihenfolge in Abhängigkeit von der Informativität variieren, um die effiziente Zielidentifikation durch den Hörer zu unterstützen. Im Gegensatz dazu erwartet die sprecherorientierte Sichtweise keinen Effekt von

Informativität auf die Reihenfolge, da Sprecher durchgängig auf bevorzugte syntaktische Strukturen zurückgreifen.

Für Forschungsfrage 2 – wie Informativität die Linearisierung moduliert – sagt die Hypothese einer informativitätsbasierten Erstpositionierung voraus, dass Sprecher die informativere Eigenschaft bevorzugt an den Anfang der Äusserung stellen, um die visuelle Suche des Hörers zu erleichtern (z.B. Cohn-Gordon et al., 2019; Fukumura, 2018; Haywood et al., 2003; Rubio-Fernández, 2016). Alternativ sagt ein auf Uniform Information Density (UID) basierender Ansatz (z.B. A. F. Frank & Jaeger, 2008; Jaeger, 2010; Jaeger & Levy, 2006) das entgegengesetzte Muster voraus: Sprecher könnten die weniger informative Eigenschaft früher platzieren, um informativitätsbedingte Spitzen zu vermeiden und die Verarbeitungslast des Hörers zu reduzieren (z.B. Hale, 2001; Sikos et al., 2017).

Forschungsfrage 3 untersucht schliesslich, ob kommunikative Einbindung den Einfluss von Informativität auf die Eigenschaftsreihenfolge moduliert. Wenn referenzielle Produktion durch hörerorientierte, effizienzgetriebene Prozesse gesteuert ist, sollten interaktive Kommunikationssettings und Aufgaben den Effekt von Informativität verstärken, da Sprecher eher die Perspektive des Hörers berücksichtigen. Umgekehrt würde ein gleichbleibendes Linearisierungsmuster unabhängig vom Grad der Kommunikationseinbindung nahelegen, dass Sprecher beim Formulieren referenzieller Ausdrücke nicht durch die Kommunikationsumgebung beeinflusst werden.

Zur Überprüfung dieser Vorhersagen wurden fünf Experimente mit einem referenziellen Kommunikationstask durchgeführt. Die experimentellen Stimuli zeigten verschiedene Tiere, die unterschiedliche Handlungen ausführten, und liessen sich im Deutschen sowohl mit prä-nominalen als auch mit post-nominalen Modifikationsstrukturen natürlich und flexibel beschreiben – wie in den Pretests in Kapitel 3 bestätigt wurde, auch wenn eine generelle Präferenz für prä-nominale Modifikationen bestand.

In den fünf Hauptexperimenten wurden jeweils zwei kritische Bedingungen miteinander verglichen: eine Animal-Informative Condition und eine Action-Informative Condition. In jeder Bedingung reduzierte die informativere Eigenschaft (gemäss RER) den Auswahlbereich von zehn auf zwei Referenten, während die weniger informative Eigenschaft ihn von zehn auf fünf reduzierte.

Die Experimente 1–3, dargestellt in Kapitel 4, untersuchten Forschungsfrage 1 und 2. Die Teilnehmer führten online einen Maze-basierten Satzergänzungstask (Maze Task") durch. Experiment 1 beinhaltete ausschliesslich den Maze Task. Experiment 2 ergänzte diesen durch einen Listener Block vor dem Maze Task, um die aktive Einbindung der Teilnehmer zu fördern. In Experiment 3 wurden zusätzlich animierte Stimuli verwendet, um eine sorgfältigere Vorschau der Szenen vor der Referenzproduktion zu unterstützen.

Über alle drei Experimente hinweg zeigten sich zwei Sprechergruppen: *Gruppe Varied* war sensitiv gegenüber Informativität und bevorzugte eine informative-first-Linearisierung, indem sie die syntaktische Struktur je nach informativerer Eigenschaft anpasste. *Gruppe Consistent* hingegen hielt im gesamten Experiment an einer einzigen syntaktischen Struktur – meist prä-nominalen Modifikationen – fest und zeigte keine Sensitivität gegenüber Informativität. Die Manipulationen

in Experiment 2 und 3 replizierten den Effekt aus Experiment 1, führten jedoch nicht zu einer Verstärkung.

Die Experimente 4 und 5, dargestellt in Kapitel 5, untersuchten Forschungsfrage 3. Diese Experimente integrierten rollenweise Perspektivwechsel in jedem Durchgang, indem Sprecher und Hörer nach jedem Trial wechselten. Experiment 4 fand wie die vorangegangenen Experimente online statt; Experiment 5 wurde als gesprochene Face-to-Face-Kommunikation im Labor durchgeführt.

Beide Experimente replizierten den Haupteffekt der Informativität. Entscheidend ist jedoch, dass in beiden Experimenten – insbesondere in Experiment 5 – ein signifikanter Anstieg des Anteils von *Gruppe Varied*-Sprechern zu beobachten war. Dies deutet darauf hin, dass kommunikative Einbindung mehr Sprecher dazu veranlasst, bei der Bestimmung der Eigenschaftsreihenfolge hörerorientierte Strategien anzuwenden. Zusätzlich wurde in beiden Experimenten ein Strukturelles Priming festgestellt (z.B. Cleland & Pickering, 2003; Pickering & Garrod, 2004; Tarenskeen et al., 2015): Sprecher wiederholten tendenziell die syntaktische Struktur ihres Partners aus dem vorherigen Durchgang – unabhängig von der Informativität.

Zusammenfassend zeigen die Ergebnisse der Experimente 1–5, dass sich zwei Sprechergruppen identifizieren lassen: Eine Gruppe (*Gruppe Consistent*) verfolgte eine sprecherorientierte Strategie, war unempfindlich gegenüber Informativität und verwendete durchgängig die bevorzugte syntaktische Struktur. Die andere Gruppe (*Gruppe Varied*) zeigte hörerorientiertes Verhalten, indem sie eine informative-first-Linearisierung bevorzugte, um die Zielsuche des Hörers zu erleichtern. Bemerkenswert ist, dass der Anteil an *Gruppe Varied*-Sprechern anstieg, wenn die Aufgabe häufige Perspektivwechsel und ein hohes Mass an kommunikativer Einbindung beinhaltete.

Kapitel 6 diskutiert diese Ergebnisse im breiteren Kontext kommunikativer Effizienz und referenzieller Produktion.

Erstens liefert die informative-first-Linearisierung robuste Evidenz für hörerorientierte kommunikative Effizienz: Informativität lenkt die Eigenschaftsreihenfolge so, dass informativere Eigenschaften früher platziert werden. Dies spiegelt das Bewusstsein des Sprechers für den inkrementellen visuellen Verarbeitungsprozess des Hörers wider und zeigt, dass Sprecher die Wortstellung nutzen, um referenzielle Unsicherheit möglichst früh zu reduzieren.

Zweitens trug das Projekt nicht nur zur Trennung sprecher- und hörerorientierter Tendenzen bei, indem zwei Sprechergruppen identifiziert wurden, sondern vereinigte diese Sichtweisen auch durch die vermittelnde Rolle der Perspektivübernahme. Perspektivübernahme ist ein flexibler Prozess (z.B. Hawkins et al., 2021), der durch kommunikative Einbindung beeinflusst werden kann. In wenig einbindenden Kommunikationsumgebungen verlassen sich Sprecher eher auf heuristische, sprecherorientierte Strategien (z.B. Koolen et al., 2013; Koolen, Krahmer, et al., 2016), was zur konsequenten Nutzung bevorzugter syntaktischer Strukturen oder universeller ÜS führt (z.B. Tourtouri et al., 2019). Im Gegensatz dazu orientieren sich Sprecher in interaktiven Tasks mit gemeinsamen Zielen zwischen Sprecher und Hörer eher an der Perspektive des Hörers und setzen Adressatenorientierung ein – in Koexistenz mit sprecherinternen Faktoren (Fukumura, 2018) – sowohl bei der Eigenschaftsauswahl als auch bei der Eigenschaftsreihenfolge.

Drittens wurde das Konzept kommunikativer Effizienz anhand der beiden Sprechergruppen neu bewertet. Das Verhalten beider Gruppen kann als effizient gelten – je nachdem, ob Effizienz auf Grundlage des Produktionsprozesses des Sprechers oder des Verstehensprozesses des Hörers definiert wird. Während informationstheoretische Ansätze und psycholinguistische Studien zur Adressatenorientierung kommunikativer Effizienz aus hörerzentrierter Perspektive betonen, sollte der anhaltende Einfluss verfügbarkeitsbasierter, sprecherorientierter Heuristiken in der Sprachproduktion nicht übersehen werden.

Zukünftige Forschung sollte die interaktive Natur von Kommunikation weiter untersuchen, etwa indem analysiert wird, wie beide Gesprächspartner ihre Strategien im Echtzeitdialog dynamisch anpassen, anstatt jeweils nur eine Rolle – Sprecher oder Hörer – isoliert zu betrachten. Die Untersuchung komplexerer syntaktischer Strukturen, wie konstituentenstruktureller Objektbeschreibungen, könnte zur Generalisierung der Informativitätseffekte auf die Linearisierung beitragen. Schliesslich könnten neuartige Kommunikationsmodalitäten, etwa human-to-computer interaction, Einblicke in kognitive Prozesse der Sprecher liefern – insbesondere solche, die mit Perspektivübernahme und Alignment mit einem Computer- oder menschlichen Adressaten verbunden sind.

Zusammenfassend untersucht diese Dissertation den Einfluss von Informativität auf die syntaktische Linearisierung der Eigenschaftsreihenfolge in der referenziellen Produktion. Durch die Integration psycholinguistischer und informationstheoretischer Ansätze leistet diese Arbeit einen Beitrag zur aktuellen Debatte zwischen sprecher- und hörerorientierten Sichtweisen der referenziellen Produktion. Anhand von fünf Experimenten zeigt das Projekt, dass Sprecher sowohl sprecher- als auch hörerorientierte Strategien bei der Linearisierung anwenden. Sprecherorientierte, heuristische Sprecher streben Produktionsökonomie an, indem sie konsistent die bevorzugte syntaktische Struktur wählen; hörerorientierte Sprecher zeigen hingegen Flexibilität in der Reihenfolgenwahl bei syntaktischer Variation. Das Linearisierungsmuster dieser Gruppe wird nicht nur von sprecherinternen Faktoren wie sprachlichen Präferenzen und Priming beeinflusst, sondern entscheidend auch von Informativität. Die informative-first-Linearisierung liefert robuste Evidenz für hörerorientierte Kommunikationseffizienz, da die frühzeitige Platzierung der informativeren Eigenschaft den Auswahlbereich möglicher Referenten für den Hörer signifikant einschränkt. Ein weiterer neuer Fund besteht darin, dass der Anteil hörerorientierter Sprecher mit zunehmender kommunikativer Einbindung steigt, was darauf hinweist, dass Perspektivübernahme eine vermittelnde Rolle dabei spielen könnte, ob Sprecher eine sprecher- oder hörerorientierte Produktionsstrategie anwenden.

Bibliography

Adamou, E., Feltgen, Q., & Padure, C. (2021). A unified approach to the study of language contact: Cross-language priming and change in adjective/noun order. *International Journal of Bilingualism*, *25*(6), 1635–1654. https://doi.org/10.1177/13670069211033909

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of hirotugu akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Alghazo, S., & Jarrah, M. (2023). No adjective ordering preferences in Jordanian Arabic grammar. *Journal of Psycholinguistic Research*, *52*(5), 1633–1667. https://doi.org/10.1007/s10936-023-09965-0

Andrade, C. (2020). Sample size and its importance in research. *Indian Journal of Psychological Medicine*, *42*(1), 102–103. https://doi.org/0.4103/IJPSYM.IJPSYM_504_19

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527. https://doi.org/10.1080/01690960801920099

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*(1), 55–70. https://doi.org/10.1016/j.jml.2004.03.006

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*(1), 361–374. https://doi.org/10.1016/j.pragma.2010.07.013

Atiega, A., Morse, S., & Elkhodairy, M. (2022). Adherence to Mirror Image Hierarchy: Evidence from Modern Standard and Libyan Arabic. *Journal of Semitic Studies*, *67*(2), 703–742. https://doi.org/10.1093/jss/fgac007

Aylett, M. P., Kristensson, P. O., Whittaker, S., & Vazquez-Alvarez, Y. (2014). None of a CHInd: Relationship counselling for HCI and speech technology. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 749–760. https://doi.org/10.1145/2559206.2578868

Aylett, M. P., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5), 3048–3058. https://doi.org/10.1121/1.2188331

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition*, *14*(3), 261–294. https://doi.org/10.1080/13506280500260484

Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during "same"-"different" decisions. *European Journal of Cognitive Psychology*, *14*(2), 237–266. https://doi.org/10.1080/09541440143000050

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.675558

Bernard, D., & Arnold, A. (2019). Cognitive interaction with virtual assistants: From philosophical foundations to illustrative examples in aeronautics. *Computers in Industry*, *107*, 33–49. https://doi.org/10.1016/j.compind.2019.01.010

Bever, T. G. (2013). The cognitive basis for linguistic structures. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path: The cognitive and biological basis for linguistic structures* (pp. 279–362). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677139.003.0001

Bezuidenhout, A. (2013). Perspective taking in conversation: A defense of speaker non-egocentricity. *Journal of Pragmatics*, *48*(1), 4–16. https://doi.org/10.1016/j.pragma.2012.11.007

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 467–484. https://doi.org/10.1016/S0022-5371(80)90321-7

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, *138*(1), 244–253. https://doi.org/10.1016/j.actpsy.2011.06.010

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25. https://doi.org/10.1016/S0010-0277(99)00081-5

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*(2), 172–189. https://doi.org/10.1016/j.lingua.2007.02.003

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Broisson, Z., & Degand, L. (2022). How egocentric is discourse marker use? Investigating the impact of speaker orientation and cognitive load on discourse marker production. In M. Cuenca & L. Degand (Eds.), *Discourse markers in interaction: From production to comprehension* (pp. 121–158). De Gruyter Mouton. https://doi.org/10.1515/9783110790351-006

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*(4), 441–472. https://doi.org/10.1016/0010-0285(87)90015-6

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900. https://doi.org/10.3758/PBR.16.5.893

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*, 11–33. https://doi.org/10.5087/dad.2011.102

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information*, *2*(2), 302–326. https://doi.org/10.3390/info2020302

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*(3), 833–843. https://doi.org/10.3758/s13423-014-0714-2

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*(4), 592–609. https://doi.org/10.1016/j.jml.2005.12.008

Byrne, B. (1979). Rules of prenominal adjective order and the interpretation of "incompatible" adjective pairs. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 73–78. https://doi.org/10.1016/S0022-5371(79)90574-7

Chartrand, T. L., & van Baaren, R. (2009). *Human mimicry* (Vol. 41, pp. 219–274). Academic Press. https://doi.org/10.1016/S0065-2601(08)00405-X

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In *Italian syntax and universal grammar* (pp. 287–309). Cambridge University Press.

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*(2), 332–373. http://www.jstor.org/stable/414102

Clark, H. H., & Marshall, C. R. (1981). Definite knowledge and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (Vol. 9, pp. 287–299). North-Holland. https://doi.org/10.1016/S0166-4115(09)60059-5

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Clarke, A. D., Elsner, M., & Rohde, H. (2013). Where's wally: The influence of visual salience on referring expression generation. *Frontiers in Psychology, Volume 4 - 2013*. https://doi.org/10.3389/fpsyg.2013.00329

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214–230. https://doi.org/10.1016/S0749-596X(03)00060-3

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404–413. http://www.jstor.org/stable/2331986

Cohn-Gordon, R., Goodman, N., & Potts, C. (2019). An incremental iterated response model of pragmatics. In G. Jarosz, M. Nelson, B. O'Connor, & J. Pater (Eds.), *Proceedings of the society for computation in linguistics (SCiL) 2019* (pp. 81–90). https://doi.org/10.7275/cprc-8x17

Cowan, B. R., & Branigan, H. P. (2015). Does voice anthropomorphism affect lexical alignment in speech-based human-computer dialogue? *Interspeech 2015*, 155–159. https://doi.org/10.21437/Interspeech.2015-75

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in humancomputer dialogue. *International Journal of Human-Computer Studies*, *83*, 27–42. https://doi.org/10.1016/j.ijhcs.2015.05.008

Cowan, B. R., Clark, L., Candello, H., & Tsai, J. (2023). Introduction to this special issue: Guiding the conversation: New theory and design perspectives for conversational user interfaces. *Human–Computer Interaction*, *38*(3-4), 159–167. https://doi.org/10.1080/07370024.2022.2161905

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3), 696–717. https://doi.org/10.1353/lan.2020.0045

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean Maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233–263. https://doi.org/10.1016/0364-0213(95)90018-7

Damen, D., van der Wijst, P., van Amelsvoort, M., & Krahmer, E. J. (2019). Perspective-taking in referential communication: Does stimulated attention to addressees' perspective influence speakers' reference production? *Journal of Psycholinguistic Research*, *48*, 257–288. https://doi.org/10.1007/s10936-018-9602-7

Davies, C., & Arnold, J. (2019). Reference and informativeness: How context shapes referential choice. In C. Cummins & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics* (pp. 474–493). Oxford University Press. https://eprints.whiterose.ac.uk/146954/

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*(8), 1956–1972. https://doi.org/10.1016/j.lingua.2010.02.005

Davies, C., & Katsos, N. (2013). Are speakers and listeners "only moderately gricean"? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*, *49*(1), 78–106. https://doi.org/10.1016/j.pragma.2013.01.004

Davies, C., & Kreysa, H. (2017). Looking at a contrast object before speaking boosts referential informativeness, but is not essential. *Acta Psychologica*, *178*, 87–99. https://doi.org/10.1016/j.actpsy.2017.06.001

Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, *9*, 519–540. https://doi.org/10.1146/annurev-linguistics-031220-010811

Degen, J., Hawkins, R., Graf, C., Kreiss, E., & Goodman, N. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*, *127*, 591–621. https://doi.org/10.1037/rev0000186

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Ding, J., Tarokh, V., & Yang, Y. (2018). Model selection techniques: An overview. *IEEE Signal Processing Magazine*, *35*(6), 16–34. https://doi.org/10.1109/MSP.2018.2867638

Dombi, J., Sydorenko, T., & Timpe-Laughlin, V. (2022). Common ground, cooperation, and recipient design in human-computer interactions. *Journal of Pragmatics*, *193*, 4–20. https://doi.org/10.1016/j.pragma.2022.03.001

Durrell, M. (2011). *Hammer's German grammar and usage* (7th ed.). Routledge.

Dyer, W. (2018). Integration complexity and the order of cosisters. In M.-C. de Marneffe, T. Lynn, & S. Schuster (Eds.), *Proceedings of the second workshop on universal dependencies (UDW 2018)* (pp. 55–65). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-6007

Dyer, W., Futrell, R., Liu, Z., & Scontras, G. (2021). Predicting cross-linguistic adjective order with information gain. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Findings of the association for computational linguistics: ACL-IJCNLP 2021* (pp. 957–967). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.83

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436. https://doi.org/10.1007/BF02143160

Elsner, M., Clarke, A., & Rohde, H. (2018). Visual complexity and its effects on referring expression generation. *Cognitive Science*, *42*(S4), 940–973. https://doi.org/10.1111/cogs.12507

Engelhardt, P. E., Alfridijanta, O., McMullon, M. E. G., & Corley, M. (2017). Speaker-versus listener-oriented disfluency: A re-examination of arguments and assumptions from autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*, 2885–2898. https://doi.org/10.1007/s10803-017-3215-0

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573. https://doi.org/10.1016/j.jml.2005.12.009

Engelhardt, P. E., Demiral, S. B., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, *77*(2), 304–314. https://doi.org/10.1016/j.bandc.2011.07.004

Engelhardt, P. E., Xiang, M., & Ferreira, F. (2008). Anticipatory eye movements mediated by word-order constraints. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*. https://escholarship.org/uc/item/8nd7f1mn

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/

10.1006/jmla.2001.2797

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 49, pp. 209–246). Academic Press. https://doi.org/10.1016/S0079-7421(08)00006-6

Ferreira, V. S. (2019). A mechanistic framework for explaining audience design in language production. *Annual Review of Psychology*, *70*, 29–51. https://doi.org/10.1146/annurev-psych-122216-011653

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. https://doi.org/10.1006/cogp.1999.0730

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*(3), 263–284. https://doi.org/10.1016/j.cognition.2004.09.002

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*(1), 163–171. https://doi.org/10.3758/BRM.41.1.163

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 939–944. https://escholarship.org/uc/item/7d08h6j4

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. https://doi.org/10.1016/j.cogpsych.2014.08.002

Fukumura, K. (2018). Ordering adjectives in referential communication. *Journal of Memory and Language*, *101*, 37–50. https://doi.org/10.1016/j.jml.2018.03.003

Fukumura, K., & Carminati, M. N. (2021). Overspecification and incremental referential processing: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(5), 680–701. https://doi.org/10.1037/xlm0001015

Fukumura, K., & Santesteban, M. (2017). *Ordering adjectives for communicative efficiency in English and Basque*. Abstract presented at the 2017 Conference on Architectures and Mechanisms in Language Processing (AMLaP 17).

Fukumura, K., & Zhang, S. (2023). The interplay between syntactic and non-syntactic structure in language production. *Journal of Memory and Language*, *128*, 104385. https://doi.org/10.1016/j.jml.2022.104385

Fussell, S. R., & Krauss, R. M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, *25*(3), 203–219. https://doi.org/10.1016/0022-1031(89)90019-X

Futrell, R. (2019). Information-theoretic locality properties of natural language. In X. Chen & R. Ferrer-i-Cancho (Eds.), *Proceedings of the first workshop on quantitative syntax (quasy, SyntaxFest 2019)* (pp. 2–15). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-7902

Gahl, S., & Garnsey, S. (2006). Knowledge of grammar includes knowledge of syntactic probabilities. *Language*, *82*(2), 405–410. http://www.jstor.or

g/stable/4490163

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51. https://doi.org/10.1016/j.jml.2009.09.002

Gann, T. M., & Barr, D. J. (2012). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience*, *29*(6), 744–760. https://doi.org/10.1080/01690965.2011.641388

Gatt, A., Goudbeek, M., & Krahmer, E. J. (2011). Attribute preference and priming in reference production: Experimental evidence and computational modeling. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 2627–2632. https://escholarship.org/uc/item/89w6j6b4

Gatt, A., Krahmer, E. J., van Deemter, K., & van Gompel, R. P. G. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, *41*(S6), 1457–1492. https://doi.org/10.1111/cogs.12375

Gatt, A., Krahmer, E. J., van Gompel, R. P. G., & van Deemter, K. (2013). Production of referring expressions: Preference trumps discrimination. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, 483–488. https://escholarship.org/uc/item/6gd9737z

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Goeke, C., Finger, H., Diekamp, D., Standvoss, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *2017 International Conference on Computational Social Science (IC2S2)*.

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, *26*(4), 300–311. https://doi.org/10.1016/j.tics.2022.01.005

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Goudbeek, M., & Krahmer, E. J. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. *Topics in Cognitive Science*, *4*(2), 269–289. https://doi.org/10.1111/j.1756-8765.2012.01186.x

Green, P., & MacLeod, C. J. (2016). Simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greenall, A. K. (2006). Maxims and flouting. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 545–548). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00361-8

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of human language* (pp. 73–113). MIT Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech Acts* (pp. 41–58). Academic Press.

Griffin, Z. M., & Bock, J. K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*(3), 313–338. https://doi.org/10.1006/jmla.1997.2547

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In E. Gibson & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference*. The MIT Press. https://doi.org/10.7551/mitpress/8957.003.0013

Gussow, A. E., & MacDonald, M. C. (2023). Utterance planning under message uncertainty: Evidence from a novel picture-naming paradigm. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(3), 957–972. https://doi.org/10.3758/s13415-023-01103-6

Hahn, M., Degen, J., Goodman, N., Jurafsky, D., & Futrell, R. (2018). An information-theoretic explanation of adjective ordering preferences. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society (CogSci)*, 1766–1771. https://cogsci.mindmodeling.org/2018/papers/0339/index.html

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, 1–8. https://doi.org/10.3115/1073336.1073357

Hartsuiker, R. J. (1999). Priming word order in sentence production. *The Quarterly Journal of Experimental Psychology Section A*, *52*(1), 129–147. https://doi.org/10.1080/713755798

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*(2), 113–157. https://doi.org/10.1006/cogp.2000.0744

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*, *45*(3), e12926. https://doi.org/10.1111/cogs.12926

Haywood, S., Pickering, M. J., & Branigan, H. P. (2003). Co-operation and coordination in the production of noun phrases. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, 533–538. https://escholarship.org/uc/item/3rw55676

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science*, *4*(2), 290–305. https://doi.org/10.1111/j.1756-8765.2012.01182.x

Hjeij, M., & Vilks, A. (2023). A brief history of heuristics: How did research on heuristics evolve? *Humanities and Social Sciences Communications*, *10*(1), 64. https://doi.org/10.1057/s41599-023-01542-z

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Huang, H.-W., & Federmeier, K. D. (2012). Dispreferred adjective orders elicit brain responses associated with lexico-semantic rather than syntactic processing. *Brain Research*, *1475*, 62–70. https://doi.org/10.1016/j.brainres.2012.07.050

Ianì, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121. https://doi.org/10.1016/j.jml.2017.05.004

Jacobs, C. L., Cho, S.-J., & Watson, D. G. (2019). Self-priming in production: Evidence for a hybrid model of syntactic priming. *Cognitive Science*, *43*(7), e12749. https://doi.org/10.1111/cogs.12749

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Jaeger, T. F., & Levy, R. (2006). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19, pp. 849–856). MIT Press. https://proceedings.neurips.cc/paper_files/paper/2006/file/c6a01432c8138d46ba39957a8250e027-Paper.pdf

Jaeger, T. F., & Tily, H. (2011). On language "utility": Processing complexity and communicative efficiency. *WIREs Cognitive Science*, *2*(3), 323–335. https://doi.org/10.1002/wcs.126

Jara-Ettinger, J., & Rubio-Fernández, P. (2022). The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychological Review*, *129*(6), 1394–1413. https://doi.org/10.1037/rev0000345

Kecskes, I. (2011). Salience in language production. In K. M. Jaszczolt & K. Allan (Eds.), *Salience and defaults in utterance processing* (pp. 81–102). De Gruyter Mouton. https://doi.org/10.1515/9783110270679.81

Kemmerer, D., Weber-Fox, C., Price, K., Zdanczyk, C., & Way, H. (2007). Big brown dog or brown big dog? An electrophysiological study of semantic constraints on prenominal adjective order. *Brain and Language*, *100*(3), 238–256. https://doi.org/10.1016/j.bandl.2005.12.002

Kempe, V., Rookes, M., & Swarbrigg, L. (2013). Speaker emotion can affect ambiguity production. *Language and Cognitive Processes*, *28*(10), 1579–1590. https://doi.org/10.1080/01690965.2012.755555

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*(2), 46–49. https://doi.org/10.1111/1467-8721.ep13175613

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110. https://doi.org/10.1207/S15327078IN0401/_5

Koolen, R. (2019). On Visually-Grounded reference production: Testing the effects of perceptual grouping and 2D/3D presentation mode. *Frontiers in Psychology*, *10*, 2247. https://doi.org/10.3389/fpsyg.2019.02247

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. J. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*(13), 3231–3250. https://doi.org/10.1016/j.pragma.2011.06.008

Koolen, R., Gatt, A., van Gompel, R. P. G., Krahmer, E. J., & van Deemter, K. (2016). Viewing time affects overspecification: Evidence for two strategies of attribute selection during reference production. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 159–164. https://escholarship.org/uc/item/34x1d0qc

Koolen, R., Goudbeek, M., & Krahmer, E. J. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, *37*(2),

395–411. https://doi.org/10.1111/cogs.12019

Koolen, R., & Krahmer, E. J. (2020). Effects of domain size during reference production in photo-realistic scenes. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, 2719–2725. https://escholarship.org/uc/item/3350n2cr

Koolen, R., Krahmer, E. J., & Swerts, M. (2016). How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science*, *40*(7), 1617–1647. https://doi.org/10.1111/cogs.12297

Koranda, M. J., Zettersten, M., & MacDonald, M. C. (2022). Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science*, *33*(9), 1440–1451. https://doi.org/10.1177/09567976221089603

Kotowski, S., & Härtl, H. (2019). How real are adjective order constraints? Multiple prenominal adjectives at the grammatical interfaces. *Linguistics*, *57*(2), 395–427. https://doi.org/10.1515/ling-2019-0005

Koulouri, T., & Lauria, S. (2009). Exploring miscommunication and collaborative behaviour in human-robot interaction. In P. Healey, R. Pieraccini, D. Byron, S. Young, & M. Purver (Eds.), *Proceedings of the SIGDIAL 2009 conference* (pp. 111–119). Association for Computational Linguistics. https://aclanthology.org/W09-3915

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development*, *40*(1), 255–266. http://www.jstor.org/stable/1127172

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American*, *236*(2), 100–105. http://www.jstor.org/stable/24953896

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1), 113–114. https://doi.org/10.3758/BF03342817

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343–346. https://doi.org/10.1037/h0023705

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kursat, L., & Degen, J. (2021). Perceptual difficulty differences predict asymmetry in redundant modification with color and material adjectives. *Proceedings of the Linguistic Society of America*, *6*(1), 676–688. https://doi.org/10.3765/plsa.v6i1.5003

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. Speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review*, *18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

Leivada, E. (2022). Determining the cognitive biases behind a viral linguistic uni-

versal: The order of multiple adjectives. *Humanities and Social Sciences Communications*, *9*(1), 436. https://doi.org/10.1057/s41599-022-01440-w

Leivada, E., & Westergaard, M. (2019). Universal linguistic hierarchies are not innately wired. Evidence from multiple adjectives. *PeerJ*, *7*, e7438. https://doi.org/10.7717/peerj.7438

Lester, M., & Beason, L. (2013). *The McGraw-hill handbook of English grammar and usage* (2nd ed.). McGraw-Hill Education.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122–142. https://doi.org/10.1037/0033-295X.98.1.122

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511813313

Levinson, S. C. (2024). *The dark matter of pragmatics: Known unknowns*. Cambridge University Press. https://doi.org/10.1017/9781009489584

Levshina, N. (2022). *Communicative efficiency: Language structure and use*. Cambridge University Press. https://doi.org/10.1017/9781108887809

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Li, C., Ferreira, V. S., & Gollan, T. H. (2022). Language control after phrasal planning: Playing whack-a-mole with language switch costs. *Journal of Memory and Language*, *126*, 104338. https://doi.org/10.1016/j.jml.2022.104338

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*(3), 550–557. https://doi.org/10.3758/BF03196312

Long, M., Moore, I., Mollica, F., & Rubio-Fernández, P. (2021). Contrast perception as a visual heuristic in the formulation of referential expressions. *Cognition*, *217*, 104879. https://doi.org/10.1016/j.cognition.2021.104879

Long, M., Rohde, H., & Rubio-Fernández, P. (2020). The pressure to communicate efficiently continues to shape language use later in life. *Scientific Reports*, *10*(1), 8214. https://doi.org/10.1038/s41598-020-64475-6

Loy, J. E., & Smith, K. (2021). Speakers align with their partner's overspecification during interaction. *Cognitive Science*, *45*(12), e13065. https://doi.org/10.1111/cogs.13065

Lyngs, U. (2019). *Oxforddown: An Oxford University thesis template for R Markdown* [GitHub repository]. GitHub. https://doi.org/10.5281/zenodo.3484681

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00226

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential

communication. *Journal of Language and Social Psychology*, *7*(3-4), 181–191. https://doi.org/10.1177/0261927X8800700403

Martin, J. E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 697–704. https://doi.org/10.1016/S0022-5371(69)80032-0

Mattmüller, U. C. (2016). *Partizipialkonstruktionen [Participle constructions]*. https://deutschegrammatik20.de/komplexer-satz/partizipialkonstruktionen-2/

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment using Kaldi. *Proc. Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Mendelsohn, M., & Straker, G. (1999). Social perspective taking and use of discounting in children's perceptions of others' helping behavior. *The Journal of Genetic Psychology*, *160*(1), 69–83. https://doi.org/10.1080/00221329909595381

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *43*(4), 907–912. https://doi.org/10.1080/14640749108400962

Mol, L., Krahmer, E. J., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human–human and human–machine interactions. *Gesture*, *9*(1), 97–126. https://doi.org/10.1075/gest.9.1.04mol

Morgan, A. M., & Ferreira, V. S. (2022). Still no evidence for audience design in syntax: Resumptive pronouns are not the exception. *Journal of Memory and Language*, *127*, 104368. https://doi.org/10.1016/j.jml.2022.104368

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive Science*, *42*(S4), 974–1008. https://doi.org/10.1111/cogs.12582

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*(4), 257–273. https://doi.org/10.1037/h0029436

OpenAI. (2025). *Image generated using ChatGPT with DALL-e (april 2025)* [AI-generated image].

Out, C., Goudbeek, M., & Krahmer, E. J. (2020). Do speaker's emotions influence their language production? Studying the influence of disgust and amusement on alignment in interactive reference. *Language Sciences*, *78*, 101255. https://doi.org/10.1016/j.langsci.2019.101255

Oviatt, S. (1995). Predicting spoken disfluencies during human–computer interaction. *Computer Speech & Language*, *9*(1), 19–35. https://doi.org/10

.1006/csla.1995.0002

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688–704. https://doi.org/10.1006/jmla.2001.2826

Paraboni, I., & van Deemter, K. (2014). Reference and the facilitation of search in spatial domains. *Language, Cognition and Neuroscience*, *29*(8), 1002–1017. https://doi.org/10.1080/01690965.2013.805796

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, *78*, 1–17. https://doi.org/10.1016/j.jml.2014.10.003

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89–110. https://doi.org/10.1515/ling.1989.27.1.89

Pechmann, T. (1994). *Sprachproduktion: Zur generierung komplexer nominalphrasen [Speech production: On generating complex noun phrases]*. VS Verlag für Sozialwissenschaften. https://books.google.de/books?id=Du1yQgAACAAJ

Pechmann, T., & Zerbst, D. (1990). Zum Einfluss kognitiver Prozesse auf die Produktion komplexer Nominalphrasen [On the influence of cognitive processes on the production of complex noun phrases]. In S. W. Felix, S. Kanngiesser, & G. Rickheit (Eds.), *Sprache und Wissen: Studien zur kognitiven Linguistik* (pp. 207–221). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-05395-8_9

Peloquin, B. N., Goodman, N. D., & Frank, M. C. (2020). The interactions of rational, pragmatic agents lead to efficient language structure and use. *Topics in Cognitive Science*, *12*(1), 433–445. https://doi.org/10.1111/tops.12489

Peña, P. R., Doyle, P., Edwards, J., Garaialde, D., Rough, D., Bleakley, A., Clark, L., Henriquez, A. T., Branigan, H., Gessinger, I., & Cowan, B. R. (2023). Audience design and egocentrism in reference production during human-computer dialogue. *International Journal of Human-Computer Studies*, *176*, 103058. https://doi.org/10.1016/j.ijhcs.2023.103058

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory and Language*, *46*(3), 586–605. https://doi.org/10.1006/jmla.2001.2824

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*(2), 97–132. https://doi.org/10.1016/S0010-0277(00)00090-1

Qing, C., & Franke, M. (2015). Variations on a bayesian theme: Comparing bayesian models of referential reasoning. In H. Zeevat & H.-C. Schmitz (Eds.), *Bayesian natural language semantics and pragmatics* (pp. 201–220). Springer International Publishing. https://doi.org/10.1007/978-3-319-17064-0_9

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rasenberg, M., Pouw, W., Özyürek, A., & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific Reports*, *12*(1), 19111. https://doi.org/10.1038/s41598-022-22883-w

Rehrig, G., Cullimore, R. A., Henderson, J. M., & Ferreira, F. (2021). When more is more: Redundant modifiers can facilitate visual search. *Cognitive Research: Principles and Implications*, *6*(1), 10. https://doi.org/10.1186/s41235-021-00275-4

Robert M., K., Robert A., D., Chen, Y., & Frances, R. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, *31*(6), 533–552. https://doi.org/10.1006/jesp.1995.1024

Rohde, H., Futrell, R., & Lucas, C. G. (2021). What's new? A comprehension bias in favor of informativity. *Cognition*, *209*, 104491. https://doi.org/10.1016/j.cognition.2020.104491

Rohde, H., Hoek, J., Keshev, M., & Franke, M. (2022). This better be interesting: A speaker's decision to speak cues listeners to expect informative content. *Open Mind*, *6*, 118–131. https://doi.org/10.1162/opmi_a_00058

Rohde, H., & Rubio-Fernández, P. (2022). Color interpretation is guided by informativity expectations, not by world knowledge about colors. *Journal of Memory and Language*, *127*, 104371. https://doi.org/10.1016/j.jml.2022.104371

Ronderos, C. R., Aparicio, H., Long, M., Shukla, V., Jara-Ettinger, J., & Rubio-Fernández, P. (2024). Perceptual, Semantic, and Pragmatic Factors Affect the Derivation of Contrastive Inferences. *Open Mind*, *8*, 1213–1227. https://doi.org/10.1162/opmi_a_00165

Rothwell, C. D., Shalin, V. L., & Romigh, G. D. (2021). Comparison of common ground models for human–computer dialogue: Evidence for audience design. *ACM Transactions on Computer-Human Interaction*, *28*(2). https://doi.org/10.1145/3410876

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*, 153. https://doi.org/10.3389/fpsyg.2016.00153

Rubio-Fernández, P. (2017). The director task: A test of theory-of-mind use or selective attention? *Psychonomic Bulletin & Review*, *24*(4), 1121–1128. https://doi.org/10.3758/s13423-016-1190-7

Rubio-Fernández, P. (2019). Overinformative speakers are cooperative: Revisiting the gricean maxim of quantity. *Cognitive Science*, *43*(11), e12797. https://doi.org/10.1111/cogs.12797

Rubio-Fernández, P. (2021). Color discriminability makes over-specification efficient: Theoretical analysis and empirical evidence. *Humanities and Social Sciences Communications*, *8*(1), 147. https://doi.org/10.1057/s41599-021-00818-6

Rubio-Fernández, P., Berke, M. D., & Jara-Ettinger, J. (2024). Tracking minds in communication. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2024.11.005

Rubio-Fernández, P., & Jara-Ettinger, J. (2020). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences*, *117*(24), 13399–13404. https://doi.org/10.1073/pnas.1922067117

Rubio-Fernández, P., Mollica, F., & Jara-Ettinger, J. (2021). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*, *150*(3), 583–594. https://doi.org/10.1037/xge0000963

Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences

approach. *Journal of Experimental Psychology: General*, *144*(5), 898–915. https://doi.org/10.1037/xge0000093

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Scontras, G. (2023). Adjective ordering across languages. *Annual Review of Linguistics*, *9*, 357–376. https://doi.org/10.1146/annurev-linguistics-030521-041835

Scontras, G., Degen, J., & Goodman, N. D. (2017). Subjectivity predicts adjective ordering preferences. *Open Mind*, *1*(1), 53–66. https://doi.org/10.1162/OPMI_a_00005

Scott, G.-J. (2002). Stacked adjectival modification and the structure of Nominal Phrases. In G. Cinque (Ed.), *Functional Structure in DP and IP: The Cartography of Syntactic Structures, Volume 1*. Oxford University Press. https://doi.org/10.1093/oso/9780195148794.003.0004

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23. https://doi.org/10.1023/A:1021928914454

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 623–656.

Shen, H., & Wang, M. (2023). Effects of social skills on lexical alignment in human-human interaction and human-computer interaction. *Computers in Human Behavior*, *143*, 107718. https://doi.org/10.1016/j.chb.2023.107718

Signorell, A. (2025). *DescTools: Tools for descriptive statistics* [R package]. https://andrisignorell.github.io/DescTools/

Sikos, L., Greenberg, C., Drenhaus, H., & Crocker, M. W. (2017). Information density of encodings: The role of syntactic variation in comprehension. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 3168–3173. https://escholarship.org/uc/item/3z24j9sx

Sikos, L., Venhuizen, N. J., Drenhaus, H., & Crocker, M. W. (2021a). Speak before you listen: Pragmatic reasoning in multi-trial language games. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, 1465–1471. https://escholarship.org/uc/item/0xc7f7wc

Sikos, L., Venhuizen, N. J., Drenhaus, H., & Crocker, M. W. (2021b). Reevaluating pragmatic reasoning in language games. *PLOS ONE*, *16*(3), 1–33. https://doi.org/10.1371/journal.pone.0248388

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. http://www.jstor.org/stable/1884852

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *22*, 161–176. https://cir.nii.ac.jp/crid/1573387450382836224

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research*, *6*(1), 1–10. http://www.jstor.org/stable/167397

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565. http://www.jstor.org/stable/1127555

Sonnenschein, S. (1982). The effects of redundant communications on listeners: When more is less. *Child Development*, *53*(3), 717–729. http://www.jstor.org/stable/1129385

Sonnenschein, S. (1984). The effects of redundant communications on listeners: Why different types may have different effects. *Journal of Psycholinguistic*

Research, *13*(2), 147–166. https://doi.org/10.1007/BF01067697

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell Publishing.

Sproat, R., & Shih, C. (1991). The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary approaches to language: Essays in honor of s.-y. kuroda* (pp. 565–593). Springer Netherlands. https://doi.org/10.1007/978-94-011-3818-5_30

Sweet, H. (1898). *A new English grammar, logical and historical: Volume 2: Syntax*. Clarendon Press.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*(3), 318–330. https://doi.org/10.1016/j.jml.2011.04.009

Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology*, *6*, 1703. https://doi.org/10.3389/fpsyg.2015.01703

Thomas, J. A. (1995). *Meaning in interaction: An introduction to pragmatics* (1st ed.). Routledge. https://doi.org/10.4324/9781315842011

Tomlin, R. S., & Myachykov, A. (2015). Attention and salience. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 31–52). De Gruyter Mouton. https://doi.org/10.1515/9783110292022-003

Tourtouri, E. N., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational overspecification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, *3*(2), 175–202. https://doi.org/10.1007/s41809-019-00032-6

Trainin, N., & Shetreet, E. (2021). It's a dotted blue big star: On adjective ordering in a post-nominal language. *Language, Cognition and Neuroscience*, *36*(3), 320–341. https://doi.org/10.1080/23273798.2020.1839664

Trainin, N., & Shetreet, E. (2023). Subjectivity predicts adjective ordering preferences in hebrew, but lexical factors matter too. *Language, Cognition and Neuroscience*, *38*(7), 1020–1026. https://doi.org/10.1080/23273798.2023.2203934

Trainin, N., & Shetreet, E. (2024). Learning speaker-specific linguistic "style" is mediated by deviance from common language use. *Language and Cognition*, *16*(1), 197–219. https://doi.org/10.1017/langcog.2023.32

Triandis, H. C., Leung, K., Villareal, M. J., & Clack, F. I. (1985). Allocentric versus idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in Personality*, *19*(4), 395–415. https://doi.org/10.1016/0092-6566(85)90008-X

Urbanek, S. (2013). *Png: Read and write PNG images* (Version 0.1-7) [R package]. CRAN. https://CRAN.R-project.org/package=png

van Gompel, R. P. G., van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review*, *126*(3), 345–373. https://doi.org/10.1037/rev0000138

van Miltenburg, E., Koolen, R., & Krahmer, E. J. (2018). Varying image description tasks: Spoken versus written descriptions. In M. Zampieri, P. Nakov, N. Ljubei, J. Tiedemann, S. Malmasi, & A. Ali (Eds.), *Proceedings of the fifth workshop on NLP for similar languages, varieties and dialects (VarDial 2018)* (pp. 88–100). Association for Computational Linguistics. https://aclant

hology.org/W18-3910/

van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, *47*(1), 100–123. https://doi.org/10.1016/j.specom.2005.06.005

Vanlangendonck, F., Willems, R. M., & Hagoort, P. (2018). Taking common ground into account: Specifying the role of the mentalizing network in communicative language production. *PLOS ONE*, *13*(10), 1–21. https://doi.org/10.1371/journal.pone.0202943

Vogels, J., Howcroft, D. M., Tourtouri, E. N., & Demberg, V. (2020). How speakers adapt object descriptions to listeners under load. *Language, Cognition and Neuroscience*, *35*(1), 78–92. https://doi.org/10.1080/23273798.2019.1648839

Waldon, B., & Degen, J. (2021). Modeling cross-linguistic production of referring expressions. In A. Ettinger, E. Pavlick, & B. Prickett (Eds.), *Proceedings of the society for computation in linguistics 2021* (pp. 206–215). Association for Computational Linguistics. https://aclanthology.org/2021.scil-1.19/

Wardlow Lane, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, *20*(4), 766–772. https://doi.org/10.3758/s13423-013-0396-1

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466–1481. https://doi.org/10.1037/a0013353

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speaker's control over leaking private information during language production. *Psychological Science*, *17*(4), 273–277. https://doi.org/0.1111/j.1467-9280.2006.01697.x

Wardlow Lane, L., & Liersch, M. J. (2012). Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes*, *27*(3), 462–473. https://doi.org/10.1080/01690965.2011.556348

Webb, A. (2023). *Mostly harmless statistics*. LibreTexts. https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Mostly_Harmless_Statistics_(Webb)/09%3A_Hypothesis_Tests_and_Confidence_Intervals_for_Two_Populations/9.03%3A_Two_Proportion_Z-Test_and_Confidence_Interval

Westbury, C. (2021). Prenominal adjective order is such a fat big deal because adjectives are ordered by likely need. *Psychonomic Bulletin & Review*, *28*(1), 122–138. https://doi.org/10.3758/s13423-020-01769-w

Westerbeek, H., Koolen, R., & Maes, A. (2015). Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00935

Whorf, B. L. (1945). Grammatical categories. *Language*, *21*(1), 1–11. http://www.jstor.org/stable/410199

Wu, S. A., & Gibson, E. (2021). Word order predicts cross-linguistic differences in the production of redundant color and number modifiers. *Cognitive Science*, *45*(1), e12934. https://doi.org/10.1111/cogs.12934

Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science*, *31*(1), 169–181. https://doi.org/10.1080/03640210709336989

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity

and adaptation in quantifier interpretation. *Journal of Memory and Language*, *87*, 128–143. https://doi.org/10.1016/j.jml.2015.08.003

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*(4), 699–707. https://doi.org/10.3758/s13423-012-0262-6

Zevakhina, N., Pasalskaya, L., & Chinkova, A. (2021). Over-specification of small cardinalities in referential communication. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.745230