Justifying the evidential use of intuitive judgements in linguistics

Karen Brøcker

PhD dissertation
January 31, 2019

Main supervisor: Samuel K. Schindler
Co-supervisor: William B. McGregor

Centre for Science Studies
Aarhus University

Abstract

Linguists use native speakers' intuitive judgements about the morphosyntactic well-formedness of sentences as a source of evidence for their grammars. This dissertation investigates the theoretical assumptions underlying this practice.

Philosopher Michael Devitt attributes a view that he calls the Voice of Competence view to generative linguists. On this view, intuitive judgements can serve as evidence for theories of grammar because their informational content is derived more or less directly from the rules of the speaker's mental grammar. The attribution of this view to generative linguists has sparked a lot of debate.

This dissertation presents the results of a questionnaire study, showing that the Voice of Competence view was not the majority view among generative linguists who participated in the study. Specifically, a majority of the generative participants rejected the idea that the content of linguistic intuitive judgements is derived from the rules of the speaker's mental grammar. The dissertation presents the view that a majority of generative participants *did* hold, as well as the majority view among non-generative participants. Participants' views on the use of experimental and informal methods respectively for collecting and analysing linguistic intuitive judgements were also surveyed.

This dissertation also presents and defends a novel account of the etiology of linguistic intuitive judgements. On this account, linguistic intuitive judgements are not based on special input from the speaker's linguistic competence nor on the speaker's theoretical concept of grammaticality (as according to another proposal in the current debate). Instead, linguistic intuitive judgements are ordinary judgements based on the speaker's immediate experience of hearing the sentence in question and their accumulated experience of hearing and evaluating sentences in this way in the past. According to this account, linguists are justified in using linguistic intuitive judgements as evidence for grammars because the speaker's experience of hearing and evaluating sentences is based on their linguistic competence, although the content of the judgement is not derived from the speaker's mental grammar. It is part of the account that linguistic intuitive judgements can be influenced by other sources as well, however, and so they may be used as a defeasible source of evidence for grammatical theories, but they cannot be naively taken as direct expressions of speakers' linguistic competence.

Resumé

I lingvistik bruges sprogbrugeres intuitive bedømmelser af, hvorvidt sætninger er morfosyntaktisk velformede eller ej, som data for grammatiske beskrivelser. Denne afhandling undersøger de teoretiske antagelser, der ligger til grund for denne praksis.

Det har vakt stor debat, at filosoffen Michael Devitt har tilskrevet generative lingvister en teori, han kalder "the Voice of Competence view". Ifølge denne teori kan lingvistiske intuitive bedømmelser benyttes som data for grammatiske beskrivelser, fordi disse bedømmelsers indhold er afledt mere eller mindre direkte af de grammatiske regler, der udgør sprogbrugerens mentale grammatik.

I denne afhandling præsenteres resultaterne af en spørgeskemaundersøgelse, der viser, at majoriteten af de generative lingvister, der deltog i undersøgelsen, ikke tilslutter sig "the Voice of Competence view". Mere specifikt afviser et flertal den antagelse, at indholdet af lingvistiske intuitive bedømmelser af sætninger er afledt af reglerne i sprogbrugerens mentale grammatik. Afhandlingen præsenterer desuden de teoretiske antagelser vedrørende lingvistiske intuitive bedømmelser, der *er* mest udbredte blandt både de generative og ikke-generative lingvister, der deltog i undersøgelsen. Deres holdninger til brugen af eksperimentelle hhv. uformelle metoder til at indsamle og analysere lingvistiske intuitive bedømmelser undersøges også.

Afhandlingen fremlægger også en ny hypotese omkring oprindelsen af lingvistiske intuitive bedømmelser. Ifølge denne hypotese er lingvistiske intuitive bedømmelser ikke baserede på ekstraordinært input fra sprogbrugeres mentale grammatik (som ifølge "the Voice of Competence view") eller på sprogbrugeres teoretiske grammatiske koncepter (hvilket er et andet forslag i den aktuelle debat). Derimod er lingvistiske intuitive bedømmelser baserede på sprogbrugeres oplevelse af at høre sætninger samt på deres tidligere erfaring med at høre og bedømme sætninger i lignende situationer. Ifølge denne hypotese kan lingvistiske intuitive bedømmelser berettiget bruges som data for grammatiske beskrivelser, fordi sprogbrugeres oplevelse af at høre sætninger er formet af deres lingvistiske kompetence, også selvom indholdet af bedømmelsen ikke formes direkte af deres mentale grammatik. Det er en del af denne hypotese, at lingvistiske intuitive bedømmelser også kan påvirkes af andet end sprogbrugeres lingvistiske kompetence. Lingvistiske intuitive bedømmelser kan derfor ifølge denne hypotese anvendes som data for grammatiske beskrivelser, men de kan ikke uden videre forbehold behandles som et direkte udtryk for sprogbrugeres lingvistiske kompetence.

Contents

Preface

In this dissertation, I focus on the intuitive judgements of morphosyntactic well-formedness that are used as evidence for grammatical theories in linguistics and on the theoretical assumptions that underlie this practice within both generative and non-generative linguistics. The dissertation draws on debates within philosophy of linguistics, different theoretical frameworks of linguistics, and formal as well as informal approaches to data collection and analysis within linguistics. Before getting started, I want to make a few remarks about my own background as well as acknowledge all the people who helped or supported me in the process of completing this project.

My background is in linguistics, more specifically in the cognitive/functional camp of linguistics. I trained as a linguist with a focus on language in use, a leaning towards experimental rather than informal approaches to data collection, and little knowledge or understanding of the motivations or practices of generative linguistics. I got my first real introduction to philosophy when starting this project. This background is likely to be reflected in this dissertation, as someone with a different background might have chosen their focus differently. In my treatment of the different linguistic frameworks it has been my aim to present things as viewed by the proponents of the respective sides.

There are many people I would like to thank for supporting me in my work to produce this dissertation. Firstly, I would like to thank Samuel Schindler who heads the research group on Intuitions in Science and Philosophy which has been the home of this project. The group is funded by the Independent Research Fund Denmark (DFF 4180-00071). As supervisor on the project, Sam has read and given valuable comments on each chapter of the dissertation, provided feedback on many versions of the questionnaire, and supplied general supervision along the way. I also owe a lot of thanks to my other supervisor, Bill McGregor, for his feedback on every chapter of the dissertation and steady guidance throughout the project. I am also hugely grateful to Anna Drożdżowicz and Pierre Saint-Germier, who gave their feedback on various stages of the questionnaire, and to Anna for discussing methodology and analysis strategies with me and for giving

me feedback on an earlier version of chapter 8 and sections of chapter 6.

I also owe many thanks to the following: to Michael Devitt for making my visit to the Philosophy Program at the Graduate Center, CUNY, in the spring term of 2018 possible, for discussing his thoughts and ideas on the intuition question with me, and for feedback on previous versions of chapter 8; to Georges Rey for discussing his views and a previous version of chapter 8 with me; to all the participants in the workshop Linguistic Intuitions, Evidence, and Expertise, Aarhus October 2017, for stimulating talks and discussions, as well as for helpful suggestions on my project; to the participants in the workshop Intuitions and the Expertise defence, Aarhus September 2017, for their insightful talks and discussions, which helped me understand the issue of intuitions in philosophy better; to the organisers of the 1st Urbino Summer School in Epistemology in 2017 for creating a great programme; to Ken Ramshøj Christensen, who discussed the generative linguistic framework with me; to Jason Kandybowicz and Sam Al Khatib who taught a course on Modern Linguistic Theories (with a focus on generative linguistics) during my stay at CUNY; and to the other professors at CUNY whose courses I sat in on, Alberto Cordero, David Papineau, and Michael Devitt, for including me in their classes and giving me greater insight in the philosophy of science and mind.

Thank you also to the many linguists who filled in my questionnaire. I highly appreciate the time and effort they put into it and the many interesting comments that many of them took the time to provide. Thanks also to the participants in the pilot test of my questionnaire, who not only sat through an early version of the study but also made time to discuss their answers and their understanding of the questions with me afterwards.

I also want to thank all my colleagues at the Centre for Science Studies at Aarhus University for being such great company throughout my three years at the centre. Thank you in particular to Randi Mosegaard for providing excellent administrative support at all stages of the project.

I received the bulk of my training as a linguist at the linguistics department at Aarhus University. A huge thank you to everyone there for providing a solid introduction to linguistics and a very friendly and stimulating environment to be a student in.

Finally, thank you to my family for all their support and encouragement, especially to Ky who moved to Denmark with me to pursue this project and patiently supported me throughout the whole process, even when I left to go to New York for four months.

Chapter 1

Introduction

Linguists commonly use native speakers' intuitive judgements about whether particular sentences are well-formed or not as evidence for their grammatical descriptions and theories.

In one introductory syntax paper, Adger (2015) presents the following set of sentences (slightly revised here):

(1) a. *The girls from Paris is singing.
b. The girls from Paris are singing.

This set of sentences, with (1a) marked as ill-formed by the prefixed asterisk (*), is used as evidence about the workings of the grammar of English. This is a fairly straightforward case that, presumably, no native speakers of English would object to. Other intuitive judgements are less clear-cut though. The examples in (2) are marked with an asterisk and a question mark (?) respectively to show that the first is ill-formed, whereas the second one is mildly ill-formed (as judged by the author of the paper in question, in this case Kayne 1983, quoted in Sprouse and Almeida 2013a, 223):

(2) a. *I'd like to know where who hid it.
b. ?I'd like to know where who hid what.

In fact, intuitive judgements about a range of different aspects of utterances are used as evidence within different fields in linguistics. In this dissertation, I will only be concerned with intuitive judgements of the morphosyntactic well-formedness of sentences, often called either "acceptability judgements" or "grammaticality judgements". In general, I will use the term "linguistic intuitive judgements" (or sometimes the shorter form "intuitive judgements") throughout the dissertation to refer to these particular judgements except when referencing

sources that use other terms, such as "acceptability judgements" or "grammaticality judgements" (in part II of the dissertation, I also use the term "syntactic intuitions" as explained in section 5.3.3).

For intuitive judgements like those in (1) and (2) to be used as evidence about the grammar of the speaker's language, we need some reason to believe that the fact that a native speaker makes the intuitive judgement that sentence S is acceptable (or unacceptable) is reliably related to S actually being acceptable (or unacceptable) in the speaker's language. The question is whether we have reason to believe that there is such a correlation for linguistic intuitive judgements?

On the face of it, using intuitive judgements, even those of experts, as evidence seems highly unscientific. Even though we might have great faith in the intuitive judgements of say, doctors, engineers, or farmers about symptoms of illnesses, the stability of bridges, or the identification of crops and weeds, we would not, strictly speaking, count those intuitive judgements as *evidence* in those respective fields. Rather, such intuitive judgements would provide a good starting point for further inquiry where more independent sorts of evidence would be required (the doctor could back up their intuitive judgement with medical tests, etc.) Is linguistics different from these other disciplines? And if so, what is it about linguistic intuitive judgements that make them fit to serve as evidence for grammatical theories?

These questions are at the centre of this dissertation. To address them, I focus on the theoretical assumptions that underlie the practice of using linguistic intuitive judgements as evidence for grammatical theories. I do this by reviewing ongoing debates from the literature about the justification for using intuitive judgements as evidence for grammatical theories and about the methodology used when collecting and analysing intuitive judgements, by surveying the views of practising linguists about these issues, and by developing and defending a novel account of how and to what extent the evidential use of linguistic intuitive judgements can be justified.

1.1 Aims

The question of why we can use intuitive judgements as evidence for grammatical theories in linguistics is a question in the methodology of linguistics as well as the philosophy of linguistics. What methodology best serves the various aims of linguistics? Do we have good reasons to suspect that intuitive judgements can serve as evidence? A general aim of this dissertation is to contribute to those general discussions.

More specifically, the main aim of this dissertation is to contribute to the debate between Michael Devitt and a number of philosophers of language and

linguistics. In that debate (introduced in chapter 2), there are two central questions. One is the normative question of what view linguists *should* hold on the question of why intuitive judgements can serve as linguistic evidence. The other is the descriptive question of what view on this question a majority of generative linguists *in fact* hold.

Devitt himself argues for a view on which linguistic intuitive judgements are every-day judgements about sentences with no special etiology, made according to the speaker's (folk) theoretical concept of "grammaticality". We are justified in using these judgements as evidence to the degree that the person making them has a good amount of experience with their language and a good (folk) linguistic theory to apply when making their judgement. He attributes a view to generative linguists in general on which intuitive judgements are the "Voice of Competence" (VoC), in that the content of the judgement is supplied by the speaker's linguistic competence. His critics instead argue for a collection of views on which competence plays a central role by supplying some special input for intuitive judgements without directly supplying the content of judgements. On both VoC and the critics' views, we are justified in using linguistic intuitive judgements as evidence because of the causal role the speaker's linguistic competence plays in the etiology of intuitive judgements.

Especially the attribution of VoC to generative linguists has been hotly debated and denied by many in the debate. One central aim of this dissertation is to bring clarity to this convoluted dispute by identifying what the majority view on this issue really is among generative linguists. As mentioned, the etiology debate between Devitt and his critics also involves the normative question of what view one should take on this issue. Another central aim of this dissertation is to contribute to that aspect of the debate by presenting and defending a novel account of the etiology of intuitive judgements which can account for the role of intuitive judgements as evidence in linguistics without invoking neither special input from speakers' competence nor speakers' theoretical concepts.

While the debate between Devitt and his critics is at the centre of this dissertation, certain other debates provide interesting perspectives for the themes and results of this dissertation as well. One such issue is the methodological debate on how to collect and analyse linguistic intuitive judgements. The experimentalist side of this debate argues that we should use formal methods, including using large samples of materials and naive subjects. The traditionalist side of the debate, on the other hand, argues that using linguists' own, informally collected judgements is sufficient. I examine the majority views on these questions among linguists to see where they fall on this issue, and I investigate potential correlations between stance on this issue and stance on the etiology debate mentioned above.

Another issue that serves as backdrop for the main theme of this dissertation is the deep theoretical divide within the field of linguistics with two separate theoretical traditions that rarely exchange results and views between them. The two main orientations within linguistics today are, broadly speaking, formal/generative linguistics and cognitive/functional linguistics (see, e.g., Newmeyer 1998, Butler 2003, ch. 1, or Darnell et al. 1999b,a, for a discussion of those terms). Tracing the historical sources and developments of this division is beyond the scope of this dissertation, but, perhaps for sociological reasons, the relationship is bitter. Differing opinions within a field of research is a good thing, as it can drive research forward, but the isolation of the two traditions of linguistics from each other is, I believe, a loss to the field of linguistics and may hamper the progress of linguistic research.

The use of intuitive judgements as evidence for theories of grammar is contentious between the two sides. As mentioned above, a central aim of this dissertation is to clarify what the majority view is on intuitive judgements used for evidence in linguistics. I will examine the majority view within both generative and non-generative linguistics, clarifying the assumptions and beliefs underlying this practice on both sides. It is my hope that this can be a small step in the direction of a better mutual understanding across theoretical boundaries. Looking into the assumptions on both sides of the divide will also give a better understanding of the particularly generative assumptions, which are at the centre of the etiology debate as mentioned above.

Finally, this dissertation may in some ways prove relevant for the debate over the use of intuitive judgements as evidence in philosophy. Some philosophers have drawn parallels between the practices of the two fields of using intuitive judgements as evidence (e.g., Machery and Stich 2013). Although the metaphilosophical debate is not the focus of this dissertation, the results of this dissertation are potentially relevant to that debate. At the end of the dissertation, I will return to that debate and consider the following questions: Will the views of linguists about linguistic intuitive judgements, their etiology, and their role as evidence be compatible with those of philosophers? Will the justification of the evidential use of linguistic intuitive judgements be applicable to philosophical intuitive judgements as well?

1.2 Intuitive judgements in linguistics

The intuitive judgements that are the focus of this dissertation are elicited by linguists when they want to investigate some grammatical phenomenon. An example could be the phenomenon of island effects and coordinate structures, where subjects could be asked to judge how acceptable the sentence “What did

they order burgers and?" sounds to them (say, in the context of another sentence saying "They ordered burgers and fries"). The linguist constructs a sentence with the relevant property and asks a native speaker whether the sentence sounds "good", "acceptable", or "natural" in their language or not (sometimes with a specific reading specified). When working on their own native language, some linguists use themselves as informants. When other native speakers than the linguist is consulted, the sentence is often, but not necessarily, presented in writing. I will only be concerned with judgements of morphosyntactic well-formedness, although other types of linguistic intuitive judgements are used as evidence within linguistics as well, including judgements of co-reference, and semantic, pragmatic, and phonological judgements.

Linguistic intuitive judgements have as their content a verdict about a sentence. This verdict may be phrased in different ways, but in order to simplify the discussion let us limit the choice to "acceptable" or "unacceptable". I only intend for these terms to represent one positive and one negative verdict for now, and we will get back to the acceptability/grammaticality distinction in chapter 2.

It is generally accepted within linguistics that those verdicts may be influenced by other factors than syntactic ones. For instance, subjects might be tired, distracted, or misunderstand the task and answer based on social acceptability rather than syntactic acceptability. It is up to the linguist to try to filter out the results that are likely to be due to grammatical differences from the ones that are likely to be due to these other factors, for instance by comparing judgements on a set of sentences that vary in relevant ways. For simplicity's sake, in my discussions of the use of linguistic intuitive judgements as evidence for grammatical theories I will be assuming that linguists are generally succeeding in this task, although this of course varies from case to case.

Even if the subject performs flawlessly, on standard generative assumptions some sentences might be judged to be acceptable by speakers though in reality be ungrammatical according to that speaker's grammar, or vice versa. The judgement of acceptability would, on this understanding, be due to what is called "performance factors" in the generative literature. We will get back to this concept below and again in chapter 2. In any case, using these intuitive judgements as evidence for theories of grammar one cannot take for granted that they are purely syntactical, and linguists in general seem to, at least in principle, recognise this.

The practice of using intuitive judgements as evidence for grammatical theories is particularly widespread in generative syntax, although it takes place within all theoretical orientations of linguistics (more on that in chapter 4). It should be noted that in all frameworks, this source of evidence is used alongside

other sources of evidence, such as corpus data, elicited production/comprehension measures from experiments, and data from other experimental methods such as brain imaging.

As mentioned above, there are currently two main theoretical orientations within linguistics. Formal/generative linguistics is the tradition inspired by the work of Noam Chomsky and others starting in the 1960s. In the generative framework, syntax is seen as autonomous from other aspects of language and language use, and it is thought to originate in a separate cognitive language module which has some innate, universal components. Cognitive/functional linguistics is a collection of traditions encompassing both European and North-American functionalism. These approaches differ on many points, but in general they share a view of language as communication and a commitment to providing general cognitive and functional accounts of language. On this perspective, language is often viewed as a skill and language acquisition as skill-learning. Further, they generally see syntax and semantics/pragmatics to be integrated rather than distinct. Butler comments on the formal vs. functional distinction and says that, "although the distinction is indeed grounded in fundamental differences of approach, it is, like most dichotomies in linguistics, over-simple and potentially misleading" (Butler, 2003, 1). Still, it gives a broad picture of the different motivations and interests within linguistics.

Because of the functionalists' focus on language as a means of communication, rather than an autonomous mental module, it is not surprising that functionalists make use of intuitive judgements as evidence for theories of grammar to a lesser degree than generative linguists do (Butler, 2006). Instead, they tend to use recorded utterances from unscripted interaction in the form of corpus data or, in line with their general cognitive commitment, data from psycholinguistic experiments.

The generative framework instead aims at modelling the speaker's linguistic competence (or, in later terminology, their I-language[1]). In generative linguistics, the term "linguistic competence" is used to refer to a speaker's abstract ability to speak a language as opposed to instances of actual language use (referred to as "linguistic performance", more on this in chapter 2). I find this general distinction useful. However, since the term is mostly used within generative linguistics, the distinction comes with certain theoretical connotations. When I use it here, I intend to use it without committing to other, usually related generative beliefs, such as the belief in a domain-specific, innate universal grammar, or the traditional generative view that only competence, and

[1]When discussing generative linguistic ideas, I will use the terms "linguistic competence", "mental grammar", and "I-language" more or less synonymously, as that is how they appear to be used in the literature. See Chomsky (1986) for the distinction between I-languages and E-languages.

not performance, should be the focus of linguistic research. Note also that the exact dividing line between what falls under competence and what falls under performance will differ with other theoretical commitments.

The generative focus on linguistic competence means that the generative linguists' higher level of use of intuitive judgements is not surprising either, since what they want to model is the implicit "knowledge" of the speaker, not language in use.

Schütze and Sprouse (2013) list some advantages of using intuitive judgements as evidence for theories of grammar from a generative point of view. They mention that intuitive judgements allow us to study structures with low frequency which, they believe, are hard to study using corpus data. They state that intuitive judgements can be used to separate impossible structures from possible structures that have never actually been produced. This kind of information, on the other hand, is not immediately obtainable using corpus data that only register occurrence or non-occurrence. Schütze and Sprouse also mention that some subjects may shows sensitivities in their intuitive judgements that they are, for some reason or another, not able or willing to produce in interaction. Finally, they mention that intuitive judgements do not include slips of the tongue and other so-called performance errors that are common in linguistic production.

This section was meant as a short introduction to the practice of using intuitive judgements as evidence for grammatical theories in linguistics and to help narrow down which linguistic intuitive judgements will be the focus throughout the dissertation. But before proceeding, let us explore the term "intuitive judgement" in a little more detail.

1.3 What are intuitive judgements?

In the philosophy literature, the terms "intuition" and "intuitive judgements" have sparked much debate (they are often used interchangeably, and I will not distinguish between them here). Bealer (1998) characterises intuitions as *a priori* intellectual seemings:

> When you have an intuition that A, it *seems* to you that A. Here 'seems' is understood, not in its use as a cautionary or "hedging" term, but in its use as a term for a genuine kind of conscious episode. (Bealer 1998, 207, original emphasis)

On this view, intuitions are a special, *sui generis* type of mental state which presents its contents as necessary and makes the subject justified in believing the content of the intuition.

On a radically different view of what intuitions are, Williamson (2007) characterises them simply as ordinary judgements about hypothetical scenarios. On his account, we use the same psychological capacities when making judgements about a hypothetical scenario as we do when we make judgements about the everyday scenarios we encounter in our day-to-day lives (though, presumably, most of the judgements that we call "intuitive" are rather immediate). On other accounts that take intuitions to be based on our everyday psychological capacities, intuitive judgements are simply thought of as beliefs (Lewis, 1983) or dispositions to believe (van Inwagen, 1997). See Pust (2017) for an overview of this literature. The main difference between the two kinds of views is that, on Bealer's view and others like it, intuitions are some one-of-a-kind mental state, and our justification for trusting these intuitions comes from their special etiology. On Williamson's view and related views, on the other hand, intuitions are underwritten by our ordinary mental capacities, and we are justified in trusting our intuitions for the reasons we trust our normal judgements, beliefs, etc. (see Machery 2011 for a discussion of these two kinds of accounts).

Regardless of the exact etiology of intuitive judgements, appeals to them in the literature seem to serve a certain purpose.[2] "Intuition talk" cannot figure in every context. The content of the intuitive judgement, it seems, cannot be directly perceptual ("I intuitively judge that this cup is black" sounds odd), and it usually cannot be explicitly inferential ("I intuitively judge that 86×9 is 774" sounds odd for most people as well, who would have to go through the calculation to verify the result). It usually also cannot be something that is to be taken on authority, and it is usually not based on overt argumentation. Rather, it seems that references to intuitive judgements in the literature function as a kind of discourse marker or dialectical manoeuvre, inviting the audience to replicate the judgement for themselves. This, of course, makes the use of intuitive judgements as evidence for some conclusion vulnerable to the risk that others do not share the intuition.[3]

[2] I owe the central ideas in this and the following paragraph to a lecture by Jennifer Nagel at the 1st Urbino Summer School in Epistemology, 2017.

[3] In this dissertation, I proceed on the assumption that intuitive judgements are in fact used as evidence for grammatical theories in linguistics. However, Santana (forthcoming) presents an alternative account on which linguistic intuitive judgements function as appeals to a shared background theory or belief within the web of scientific claims (Quine, 1951). Not all scientific claims or ideas can be debated at the same time, some have to be kept constant as part of the background while other ideas are in focus. Santana argues that judgements such as "the sentence 'the cat is on the mat' is acceptable" is not used as evidence that this sentence is, in fact, acceptable in the speaker's language, it is rather an appeal to the presumed shared background belief that this sentence (type) is acceptable. Here, disagreement would be a signal that what was assumed to be a shared background belief is, in fact, not shared and, perhaps, needs to move into the focus of current investigation. While I think this is a plausible account of at least some of the "intuition talk" that occurs in linguistics, I will proceed on the assumption that, at least sometimes, intuitive judgements are used as evidence for grammatical claims and theories (think, for instance, of intuitive judgements collected as part of fieldwork).

As for the difference between intuitive and reflective judgements, it seems likely to me that one can make a pre-theoretical, processing-based distinction between intuitive judgements on the one hand and reflective judgements on the other hand. This will, however, not be a hard-and-fast distinction. Intuitive judgements are characterised by being entrenched and automatic, but otherwise they differ little from reflective judgements. The difference between intuitive and reflective judgements, on this distinction, are available to consciousness (we can tell whether a judgement felt intuitive or reflective to us), and so, even in the absence of a fully developed theory of intuitions, one can talk about the two kinds of judgements.

As chapter 2 will show, linguists have made similar proposals to those from philosophy about intuitive judgements, with one side proposing that linguistic intuitive judgements have a special etiology, and the other side arguing that they are every-day judgements about sentences. For now, I do not intend to take a stance on this issue. Instead, for the moment I use "intuitive judgement" as a way to refer to the cluster of intuitive judgements of morphosyntactic well-formedness of sentences which are used as evidence for grammatical theories in linguistics, however their type or etiology is best characterised. This question will the in focus of later parts of the dissertation.

1.4 Overview of the dissertation

This dissertation is at the intersection between linguistics and philosophy, and it deals with both generative and non-generative approaches to linguistics, as well as with both experimental and informal approaches to data collection and analysis. I hope this dissertation will be of interest to both generative and non-generative linguists, philosophers, and to proponents of both informal and formal data collection methods. Although certain readers might find some sections to be overly detailed and others fairly brief (or vice versa), I have intended to strike a balance so that the exposition in this dissertation will be accessible to all these groups of readers.

The dissertation is organised into three parts. Part I is the background section containing the literature review. Part II consists of chapters describing the empirical study and discussing its results. Part III contains the development of my proposal for a justification for the use of intuitive judgements as evidence for grammatical theories in linguistics. I present each section in more detail below.

The background section is comprised of chapters 2, 3, and 4. In chapter 2, I lay out the etiology debate between Devitt and his critics as mentioned above. I present the three main views of the debate and dissect them into component

parts that later form the basis for some of the questions in the questionnaire presented in part II of the dissertation.

In chapter 3, I review the debate between the proponents of so-called Experimental Syntax and proponents of the traditional armchair method of collecting and analysing linguistic intuitive judgements. I also briefly comment on how this debate relates to the etiology debate discussed in chapter 2.

Having dealt with these two aspects of the use of intuitive judgements as evidence for grammatical theories from a mainly generative angle, I then turn to non-generative linguistics in chapter 4. Since there is much less debate on this in non-generative linguistics than in generative linguistics, this chapter is fairly short. I review how linguistic intuitive judgements are used in fieldwork across different theoretical orientations. I then present two non-generative accounts of the etiology of linguistic intuitive judgements as well as guidelines from the non-generative literature for the use of intuitive judgements as evidence for grammatical theories.

Chapters 5, 6, and 7 make up the empirical part of the dissertation. In chapter 5, I present the design of the study as well as the questionnaire itself. The study was performed online and was sent out to linguists using LinguistList. In the call for participants, researchers of any theoretical orientation who thought that intuitive judgements could in some way serve as evidence in linguistics were invited to participate. The questionnaire contained twelve questions based on the debates laid out in chapters 2 and 3 and a number of background questions as well.

Chapter 6 presents the results of the study over several subsections.[4] First, I review the results pertaining to the generative linguists' view on the etiology question. These results show that the majority of generative participants did not subscribe to any of the views put forward in the etiology debate. They did, however, subscribe to a view on which our intuitive judgements are good evidence for theories of grammar because of their causal connection to the speaker's competence, although when combining the majority answers to all questions, the resulting view is not the same as either of the competence-based views presented in the debate. I then present the results regarding the generative participants' views on methodology. Here, we do not see a majority for either the experimentalist or traditionalist side. Then I review the non-generative participants' views on the etiology question. The majority view in this group is more similar to the view proposed by Devitt than to the competence-based views in the debate. The non-generative participants were also found to be more experimentalist than traditionalist. I then present the results of a number of association

[4]Tables with all analysis results can be found in the appendices. The data files (anonymised) can be sent upon request.

analyses, investigating whether there were any significant associations between theoretical orientation, linguistic specialisation, and stance on the methodology question respectively on the one hand and answers to individual questions on the other hand. One of the results from these analyses was that generative linguists were significantly more likely to subscribe to some of the views associated with VoC than non-generative linguists were. Finally, in chapter 7, I summarise the results presented in chapter 6 and discuss them.

The justification part of the dissertation has one chapter, chapter 8. Here, I develop and defend an account of the etiology of linguistic intuitive judgements. On this account, linguistic intuitive judgements are every-day judgements about sentences, based on the speaker's immediate experience of hearing the sentence. According to this account, the ability that we all have as native speakers to recognise whether a sentence is part of our language or not shapes the input for the speaker's judgement. I propose that evidence from conversational data suggest that this ability is, at least sometimes, available as the basis for conscious, verbal reactions. The judgement itself is not, on this account, based on the speaker's theoretical concepts (at least not when all goes well). Instead, it is based on the speaker's accumulated experience with "hearing" whether a sentence is part of their language or not. Since we do not have introspective access to the source of our judgements, however, this account also allows that other distinctions we are sensitive to might interfere and create noise. All in all, on the account that I present, linguistic intuitive judgements are a defeasible source of evidence which can be influenced by noise from a number of different directions that we have to be aware of.

In chapter 9, I summarise and discuss the main conclusions from the previous chapters and return to the broader perspectives on the main theme of the dissertation introduced above.

Part I

Background

Chapter 2

The etiology debate

In this and the following chapter, I summarise two discussions within generative linguistics and the philosophy of linguistics respectively about the use of linguistic intuitive judgements as evidence for grammatical theories. First, the discussion in the philosophy of linguistics of how speakers' mental grammars contribute to the content of linguistic intuitive judgements, and second the discussion within generative linguistics of how to collect and analyse linguistic intuitive judgements.

Intuitive judgements are a very prominent source of evidence in generative syntax. However, intuitive judgements are used as evidence for grammatical theories outside generative linguistics as well. In chapter 4, I review what role linguistic intuitive judgements play in other theoretical frameworks of linguistics than the generative one.[1]

2.1 The justification question

Despite being a widely used source of evidence in generative syntax, there is little discussion within generative linguistics of what makes intuitive judgements good evidence for grammatical theories. To what extent and why are linguists justified in relying on intuitive judgements as evidence for their theories of grammar? Let us call this *the justification question*.

When attempting to justify the evidential use of linguistic intuitive judgements, authors typically take one of three approaches (Santana, forthcoming). The first one is to point to the fruitfulness of the practice (see, e.g., Gross and Culbertson 2011). If relying on intuitive judgements can be shown to be fruitful,

[1] In accordance with the Graduate School of Science and Technology rules, parts of chapters 2, 3, and 4 are based on sections of the progress report for the qualifying examination (more specifically, sections 2.2, 3.1, 3.2, 3.3, 3.4, 4.1, and 4.2.2 of the dissertation). These sections have all been updated and expanded.

then surely the practice is justified, even if we cannot (yet) say what makes it fruitful. The second approach to justifying the practice is to point to convergence of this method with other methods (see, e.g., Maynes 2012a). Again, we might not at present be able to say what makes linguistic intuitive judgements good evidence for theories of grammar, but if we can show that this type of evidence converges with other sources which we have good reason to trust, then we can surely use intuitive judgements as evidence without worrying too much about why they can serve as evidence. Finally, the third approach is to present an etiology story that explains why linguistic intuitive judgements can serve as evidence for grammatical theories. On this approach, the origin of linguistic intuitive judgements is used to argue for their ability to serve as evidence.

It is this third approach to justifying the evidential use of linguistic intuitive judgements I will focus on here. Let us call the question of what the etiology of intuitive judgements is *the etiology question*. Such an etiology story is what is known as a linking hypothesis in cognitive science.[2] A linking hypothesis is a hypothesis explaining how the behaviour we observe (such as eye movement patterns) could be linked to underlying cognitive mechanisms. In this case, we want a hypothesis (or etiology story) that links the overt behaviour of responding to sentences with an intuitive judgement to the underlying cognitive workings of the mind.

The particular debate I will focus on in this chapter began with the publication of *Ignorance of Language* (Devitt, 2006b,c). Since then, Devitt and Rey (2013), Maynes and Gross (2013), Ludlow (2011), and Collins (2008), among others, have been in an at times intense discussion on the topic of the etiology of linguistic intuitive judgements. Their debate has two main themes: 1) how can the etiology of intuitive judgements justify their evidential use in linguistics, and 2) what is the received view on the etiology question among generative linguists.

Before getting to the main views of the debate, I will first introduce the two main mental systems that are commonly mentioned in the debate as being relevant to the etiology of linguistic intuitive judgements. One is the speaker's linguistic competence. In generative linguistics, the term "linguistic competence" is used to refer to a speaker's abstract ability to speak a language as opposed to instances of actual language use (referred to as "linguistic performance", more on this in section 2.3.4).

The other system that is often thought to play a role in forming intuitive linguistic judgements is the meta-cognitive system which we can call the cen-

[2]This point is due to Jon Sprouse, who made a comment to this effect at the workshop Linguistic Intuitions, Evidence, and Expertise in October 2017 in Aarhus, Denmark. See also Sprouse (2018).

tral processor, where judgements, decisions, beliefs etc. are formed. The central processor is a metaphor used on computational-representational, fodorian accounts of cognition (Fodor 1983, or see Bermúdez 2010, ch. 10, for a brief introduction). On these accounts, the central processor is the meta-cognitive, non-encapsulated, domain-general central system that receives input from modular, encapsulated, domain-specific systems such as the perception module and language module. On these accounts, processes such as practical reasoning and belief fixation take place in the central processor, and this is where propositional content such as linguistic intuitive judgements would be formed. To simplify the discussion, I will use this metaphor of one central, general-purpose processor. The neuro-cognitive reality might be more complex than that (see discussions in Donald 1991, 54-58, and in Kieras 2007), but for our present purposes this abstraction will do.

2.2 Introducing the main views

2.2.1 The Voice of Competence view

The publication of Devitt (2006b) launched a round of discussions over how to view the etiology of the intuitive judgements used as evidence in linguistics (and in philosophy as well, see, e.g., the debate between Devitt 2011, Machery et al. 2013, and Devitt 2012). In this and subsequent works, Devitt characterises what he takes to be the received view on the justification question among generative linguists. He calls this view the Voice of Competence view (VoC).[3]

On this view, as Devitt characterises it, we are justified in relying on linguistic intuitive judgements as evidence for theories of grammar because they are fairly directly caused by the speaker's linguistic competence. More specifically, the speaker's linguistic competence provides the *informational content* of the judgement without input from the central processor. A core part of the account is the assumption that the rules of a speaker's grammar are represented in that speaker's mind/brain. Then, when the speaker gets presented with a sentence and is asked to make an intuitive judgement about that sentence, the speaker unconsciously derives an answer from the rules that are represented in their mind. If the stimulus sentence is not permitted by the rules, the content of the intuitive judgement will be "not acceptable", and if the sentence is permitted by the rules, the content will be "acceptable".[4] In this way, the speaker's linguistic

[3]Although others have used the term "the Voice of Competence" to refer to somewhat different views on the etiology question (Miščević 2006, Textor 2009, Rey 2013), in the following, I use "VoC" specifically to refer to the view as it is characterised by Devitt.

[4]For now, I gloss over whether ordinary speakers can give acceptability judgements or instead give grammaticality judgements. See section 2.3.2 below.

competence provides the content of their intuitive linguistic judgements. On this view, there is still room for noise from performance factors, such as lack of attention, limitations on working memory, etc., to influence intuitive judgements in the same way that there is room for noise when the competence is applied in language production or comprehension.

Devitt (2006c, 482) criticises VoC for being cognitively immodest, since it requires that speakers "[stand] in an unconscious or tacit propositional attitude" to the rules or principles of their language and that these are represented in speakers' minds.

Devitt (2006c) notes that what he calls the "standard version" of VoC, in which rules are represented in speakers' minds, might not be the received generative view. Instead, he suggests, generative linguists might believe in a "non-standard" version of VoC in which the speaker's competence still supplies the informational content of their intuitive judgements but where this is based on embodied rather than represented rules. On my reading of Devitt's characterisation of VoC, even on the non-standard version of VoC there is no involvement of the central processor. Devitt thinks that no plausible story of how embodied rules could issue in the content of an intuitive judgement without involvement of the central processor is likely to be forthcoming, and so he rejects the non-standard version of VoC as well.

Although some use the phrase "the Voice of Competence" when arguing for somewhat different competence-based views, so far no one has come forward to defend this *particular* version of a competence-based view, to my knowledge (neither the standard nor the non-standard version).

VoC seems to follow naturally from traditional Chomskyan, mentalist views on which talk of representations, rules, and knowledge feature heavily. Devitt (2006c) lists a couple of quotes that support his reading, including the following:

> it seems reasonably clear, both in principle and in many specific cases, how unconscious knowledge issues in conscious knowledge ... it follows by computations similar to straight deduction. (Chomsky 1986, quoted in Devitt 2006c, 483)

Quotes like this on do seem to invite the conclusion that VoC is (or was, at least at some point) the received view within generative linguistics. However, perhaps the main point of criticism against Devitt (2006c) has been that his attribution of VoC to generative linguists is incorrect (more on this in the following section).

2.2.2 The Critics' Views

A number of philosophers of language and linguistics working within the tradition of generative linguistics (in the following simply "Devitt's critics") disagree with Devitt's claim that VoC is the received view among generative linguists. Different critics focus on different aspects of VoC in their responses to Devitt, but the views they present as alternatives to the VoC account do share some characteristics as well. They typically object to the idea that competence delivers the full informational content of intuitive judgements. They also object to Devitt's own view (presented below) that the intuitive judgements used as evidence in linguistics are the theoretical judgements or opinions (or, as some put it, "hunches"; Fiengo 2003) of speakers. Instead, they often compare the linguistic intuitive judgements that are used as evidence for grammatical theories to subjects' reports of their perception of stimuli in other fields of study, i.e., a subject's report of how bright a flash of light appeared to them in a psychophysics experiment.

In the following, I will treat this group of responses to Devitt's view under one heading and simply refer to them as "the Critics' Views". This is a bit of an oversimplification. The reason they are presented under one heading here is to create an overview of the aspects of the discussion where critics of Devitt have advanced alternatives to VoC and Devitt's own account. For a full presentation of each of Devitt's critics' views, I refer to their publications as listed below.

For my characterisation of the Critics' Views, I have used the following responses to Devitt's characterisation of VoC: Miščević (2006), Collins (2008), Pietroski (2008), Culbertson and Gross (2009), Gross and Culbertson (2011), Fitzgerald (2010), Maynes and Gross (2013), Rey (2013, forthcoming), and Textor (2009). Although some of these do not explicitly argue *for* what I am calling the Critics' Views, I do think they contribute to the debate by clarifying what they take to be the received view within generative linguistics.[5] I will note diverging opinions that I am aware of.

Devitt's critics think that VoC does not exhaust the possibilities of a competence-based view. They do agree with VoC that the speaker's competence plays a central role in the production of the intuitive judgement and that this is what makes intuitive judgements good evidence for linguistic theories. However, they do not think the speaker's competence provides the informational content of the judgement. Instead they think that the competence is responsible for producing some special input which is then fairly directly translated into the content of an

[5]Ludlow (2011) is excluded from this list, although he has certainly criticised Devitt (2006c), as I understand him to differ from Devitt mainly in his stance on whether or not VoC is the main view on the justification question in generative grammar. The view that Ludlow proposes himself is, in my understanding, very much like Devitt's own account.

intuitive judgement by the central processor.

Devitt (2013), however, criticises this idea, arguing, in effect, that if the role of the competence is only to supply data for the judgement, then views like these are basically just different versions of his own account (presented below) since "the intuitions arise from the subject's central processor reflection on the data of trying to understand the string". But on at least some of the critics' views, the competence provides special input for the central processor, which, as we shall see, is not the case on Devitt's account.

2.2.3 The Modest Explanation

Devitt (2006c) also presents and defends his own account, which Textor (2009) dubs the Modest Explanation, reflecting that, unlike VoC, it only relies on "cognitive states and processes that we are *already* committed to" (Devitt 2006c, 496, original emphasis). On the Modest Explanation, a speaker's linguistic intuitive judgements are likely to be good evidence for grammatical theories because a native speaker of a language has ready access to a lot of linguistic data and therefore has a relatively good empirically based theory of which utterances are part of their language and which ones are not (of course, speakers may be wrong, but when they are right, it is for this reason).[6]

It is perhaps wrong to say that Devitt's account is an attempt to answer the justification question by way of an etiology story. On his account, it is not any special etiology of the linguistic intuitive judgements that justifies their use as evidence, rather it is the degree to which the theory involved in the judgement is dependable and the subject having the right experience with reflecting on their language that justifies the judgement being used as evidence. His account is thus an alternative to the etiology-based accounts discussed above. Still, it is included here as the etiology story it involves differs from the other ones considered, even if the justification for the use of linguistic intuitive judgements as evidence does not come from the etiology of the judgements.

On the Modest Explanation, it is the central processor rather than the speaker's competence that plays the central role. The subject hears a sentence and simulates the experience of attempting to comprehend or produce the sentence. Then, using themselves as a guide to what a competent speaker would do, they use that experience as the datum to base a judgement on. At least, this is the case for difficult cases, whereas simpler cases might be decided without any direct reflection on a simulation (Devitt, 2010b). The resulting

[6] On this view, an intuitive judgement is a sort of meta-judgement, a judgement about what one would judge acceptable if one were to encounter it. Devitt argues that, for this reason, we should limit the use of these judgements as evidence for theories of grammar and instead use more direct evidence, for instance elicited language use.

intuitive judgements are "*fairly immediate unreflective judgments*" which are "*judgments about* linguistic performances" (Devitt 2006c, 482, original emphasis). Such meta-linguistic judgements are like any other judgements made by a person, except that they are produced fairly unreflectively and immediately, say, like the fairly immediate and unreflective judgement by a palaeontologist of whether an object found in the ground is made of bone or not (Devitt, 2010b).

Contrary to the competence-centred accounts, there is no special input from the speaker's linguistic competence to their central processor on this account. If the subject finds the sentence okay, they will likely judge it to be grammatical, and if not they might judge it to be ungrammatical or maybe grammatical but infelicitous. The judgement will be theory-laden, not in the sense of being based on theoretical reasoning, but rather in the sense that all judgements are theory-laden. This includes the theory-ladenness of observation judgements of the type "the grass is brown" (Devitt, 2006c). Devitt writes:

> The anti-positivist revolution in the philosophy of science, led by Thomas Kuhn and Paul Feyerabend, drew our attention to the way in which even the most straightforward judgments arising from observational experiences may depend on background expertise. We would not make the judgments if we did not hold certain beliefs or theories, some involving the concepts deployed in the judgments. We would not make the judgments if we did not have certain predispositions, some innate but many acquired in training, to respond selectively to experiences. (Devitt, 2015, 38)

But the theory-ladenness of intuitive judgements also includes the possibility that intuitive judgements are laden with folk or expert theories. Devitt mentions the potential influence from "folk linguistics, partly reflecting 'the linguistic wisdom of the ages'", and folk linguistic concepts (Devitt, 2006c, 498-499).[7] Similarly, the linguistic theories that a linguist knows are likely to influence their intuitive judgements as well. Devitt notes that:

> [t]he linguists' skill at identifying items with and without a syntactic property like, say, the biologist's skill at identifying items with and without a biological property, is likely to be better than the folk's because their theories are better. (Devitt, 2006c, 499)

Linguistic intuitive judgements of naive speakers are theory-laden in the way all observational judgements are, and one part of this theory-ladenness is that the concepts they employ about their linguistic world are the concepts of

[7]Folk linguistics is not meant to refer to prescriptivist notions of grammar like the rules for correct grammar that one might learn in school (Devitt 2010b, 254; 2010a, 841, 857).

folk linguistics. The intuitive judgements of experts are similarly theory-laden, but their training and expert knowledge has given them both more exposure to relevant cases and a different set of (hopefully) more refined theories and concepts to apply. In simple cases, however, the judgements of both groups might be more like the simple observational judgements mentioned above ("the grass is brown" etc.).

One consequence of the Modest Explanation is that, on this view, we should prefer the linguistic intuitive judgements of experts over those of lay subjects, as experts' intuitive judgements are laden with our best current linguistic theories while the intuitive judgements of lay subjects are laden with folk linguistics. To Devitt's critics, this is a problem for the Modest Explanation. On their view, we should not expect experts to be better subjects than lay people, since expertise in linguistics is not likely to improve one's linguistic competence in one's native language (see, e.g., Culbertson and Gross 2009 and Rey forthcoming). They also reject the idea that linguistic intuitive judgements are based on subjects' theoretical beliefs (folk or otherwise) about grammaticality.

The above sections were intended as a brief introduction to each of the three views that are central in the etiology debate. In the rest of this chapter, I go further into the main areas of disagreement that I think are critical to the debate. These form the basis of the questionnaire presented in part II of the dissertation.

2.3 Disagreements

The seven areas of disagreement that I will focus on in the rest of this chapter are the following:

1. Are linguistic intuitive judgements useful as evidence because of their connection to competence or because of their connection to the speaker's experience with reflecting on their language?

2. Should we use judgements of acceptability or grammaticality as evidence for grammatical theories?

3. Does the speaker's linguistic competence supply the data for intuitive judgements or does it supply the full informational content?

4. Do speakers have direct, Cartesian access to truths about their language through linguistic intuitive judgements (noise from performance factors aside)?

5. Should we have a mentalist or non-mentalist conception of linguistics and grammatical research?

6. Should we think that the structure rules of a speaker's language are somehow implemented in their mind? Or is it enough that whatever is in the speaker's mind respects those rules?

7. And if the structure rules are implemented in speakers' minds, are they represented or merely embodied?

Below, I will go through each of the identified areas of disagreement and lay out what position is taken on each of the three main views of the debate. See an overview of the respective positions in table 2.1 (which will be presented again in the summary in section 2.4 at the end of the chapter).

Variables / Views	Modest Explanation	VoC	Critics' Views
Origin: experience or competence	Experience	Competence	Competence
Acceptability or grammaticality	Grammaticality	Acceptability	Acceptability
Content or data from competence	Data	Content	Data
Direct access	No	Yes	-
Conception of linguistics	Non-mentalist	Mentalist	Mentalist
Implemented structure rules	No	Yes	-

Table 2.1: The three views

2.3.1 Origin of intuitive judgements

One fundamental disagreement between Devitt and his critics is over why linguistic intuitive judgements can serve as evidence for linguistic theories. On the Modest Explanation, the main reason linguistic intuitive judgements can serve as evidence (to the extent that they can) is because of their relation to our empirically acquired, meta-linguistic expertise with the language(s) we speak. On both VoC and the Critics' Views, the main reason linguistic intuitive judgements can be used as evidence is because they have a close, causal connection to speakers' linguistic competence.

The Modest Explanation: On the Modest Explanation, when linguistic intuitive judgements can tell us something about a speaker's language it is because they are made by a speaker who has a relatively good theory of what sentences are part of their language based on empirical experience with their language. Note that this is independent of whether our competence with our native language is acquired empirically rather than being innate. It is the view that our meta-linguistic *judgements* are based on empirical experience.

On this account, speakers' intuitive judgements are also influenced by the linguistic theories that speakers subscribe to, whether those are folk linguistic

theories or particular theories from the linguistic literature (see more on this theory-ladenness in section 2.2.3 above).

VoC and the Critics' Views: On both VoC and the Critics' Views, linguistic intuitive judgements can serve as evidence for theories of grammar because they, at least partly, are a product of the speaker's linguistic competence. Fitzgerald (2010, 130, 144), for instance, claims that competence is "responsible for" intuitive judgements and "shapes" them. Similarly, Miščević (2006, 526) outlines the view that what happens when a speaker delivers a linguistic intuitive judgement is that their cognitive apparatus "is mobilizing the particular competence, i.e. the same cognitive resource that produces or fails to produce similar sentences in real-life speaking."

Devitt takes this as evidence that his critics in fact endorse VoC, but the critics deny it. The difference may lie in how one specifies the connection between competence and linguistic intuitive judgements, more on this in section 2.3.3.

Whether one adopts the Modest Explanation, VoC, or the Critics' Views has consequences for what type of subjects one should prefer to collect linguistic intuitive judgements from. It follows from the Modest Explanation that the subjects with the most relevant experience and theories are more relevant to collect intuitive judgements from than subjects with less relevant experience and theories. In other words, it is more relevant to ask experts who have a lot of experience with linguistic data and a lot of practice producing this kind of linguistic intuitive judgements and who furthermore have refined theoretical concepts at their disposal than to ask naive subjects who, although they as native speakers have a lot of experience with their language, might have less practice and experience with linguistic intuitive judgements and who will rely on folk-linguistic concepts. Note, though, that being a linguist does not make you an expert on all and any area of linguistics, and only in the areas where you have expert knowledge are you to be preferred over other subjects (Devitt, 2010a). In contrast, on VoC and the Critics' Views, all speakers are, at least in principle, equally good subjects to elicit intuitive judgements from.

2.3.2 Acceptability or grammaticality

Another much debated disagreement between Devitt and his critics is over whether the linguistic intuitive judgements used as evidence for theories of grammar by linguists should be grammaticality or acceptability judgements. In the following, I first lay out how these terms are used before proceeding to outline the different views.

In the generative linguistic literature, the linguistic intuitive judgements in question here are sometimes referred to as *acceptability* judgements and sometimes as *grammaticality* judgements. The distinction between the two terms in this context goes back to Chomsky (1965), who defines a sentence as acceptable to a speaker if that speaker finds the sentence "natural and immediately comprehensible [...] in no way bizarre or outlandish" (Chomsky, 1965, 10). Note that this specific use of the term "acceptable" differs from more everyday uses of the term as well as from the use of the term in other areas of linguistics like pragmatics, where it might mean socially acceptable, felicitous, or appropriate to a speaker in a certain context. For a sentence to be grammatical, in this tradition, means that the structure of the sentence is in accordance with the (mental) generative rules of the language (Chomsky, 1965). It is frequently noted that the term "grammaticality judgement" is widely used in the generative literature to refer to what should, by Chomsky's definition above, be called "acceptability judgements", perhaps adding to the confusion over what the actual practice of generative linguists is (Schütze, 2006/1996; Den Dikken et al., 2007; Culbertson and Gross, 2009).

On the traditional generative view, an acceptability judgement is a speaker's report of how acceptable, natural or good a sentence strikes them to be. This is seen as a pre-theoretical or quasi atheoretical judgement (Maynes, 2012b). Acceptability, on this view, is related to grammaticality in that a sentence's status as actually being grammatical or not with respect to the mental grammar of a speaker is a crucial part of the reason why a sentence will be experienced as being (un)acceptable by that speaker. But other things might influence this experience as well, such as whether the sentence is highly taxing for the working memory. This, in the traditional generative view, is the reason that the sentence "the dog that the woman that the girl knew saw ran" will be judged as unacceptable by most speakers. On this view, the sentence is hypothesised to actually be grammatical, i.e., to be in accordance with speakers' mental grammars, but the sentence will still be judged to be unacceptable because it overworks the speakers' working memory.

In the generative literature, you find two different uses of the term "grammaticality judgement". On the one hand, "grammaticality" gets interpreted as meaning something like actually "generated by the speaker's [mental] grammar" (Schütze, 2006/1996, 20). In those contexts, generative linguists will usually remark that speakers do not have such intuitive judgements, that is, speakers do not have intuitive judgements about whether a sentence is actually permitted by their mental grammar or not (see, e.g., Schütze 2006/1996, 26). On the generative view, the closest thing that speakers have is acceptability judgements.

In other cases, "grammaticality" means something like "explanatory hypothe-

ses based on whatever relevant data and theory linguists might possess" about "what sentences are permitted, or generated, by a grammar" (Culbertson and Gross, 2009, 722). On this view, if a sentence is judged by some speaker to be grammatical, this means that the speaker has compared the sentence to their specific concept of what it means for something to be grammatical and found that the sentence is in accordance with that concept (here, "grammar" and "grammatical" refers to a specific *hypothesis* about the mental system, not the mental system itself). On this interpretation, whether the sentence is deemed to be grammatical or not depends on the person's theory of grammar. Saying that a sentence is grammatical in this sense is thus, for generative linguists, a hypothesis about how the actual mental grammar of a speaker is organised, i.e., it is a hypothesis that this sentence is permitted by the speaker's actual mental grammar. Likewise, saying that a sentence is ungrammatical would be to hypothesise that this sentence is not permitted by the speaker's actual mental grammar. In this sense, grammaticality judgements are not usually considered evidence in the generative literature. Myers (2009b, 412) writes: "a grammaticality judgment, strictly speaking, is a theoretical claim, not evidence at all".

Now, let us look at each of the three main views from the etiology debate on the acceptability/grammaticality issue.

The Modest Explanation: On the Modest Explanation, the relevant sort of linguistic intuitive judgements for linguists to build their theories on is *grammaticality* judgements[8] (when Devitt talks about grammaticality judgements, he is talking about the second sense discussed above, where "grammaticality" means according to a theory of grammar). He thinks that lay subjects will often interpret "acceptability" to mean grammatical in their folk sense of that term which might be the closest applicable concept they have (see, e.g., the response of Devitt 2010a to Culbertson and Gross 2009). According to Devitt, it is only interesting or relevant for grammatical theories to collect speakers' acceptability judgements to the extent that the speakers interpret the term "acceptability" to mean their concept of "grammaticality". This, Devitt (2010a, 843) writes, is because that while acceptability in the generative sense of the word might explain *why* or *when* speakers choose to judge a sentence to be acceptable or not, the *content* of the term "acceptability" for lay speakers will not be acceptability in the generative sense. If it were indeed possible to collect *actual* acceptability judgements that were not about grammaticality, these would not be the kind of evidence linguists should rely on for their theories of grammar

[8]Although Devitt notes that grammaticality judgements should only be a complement to other types of data like corpus data, elicited data, or acquisition data (Devitt, 2006c, 486). More on this in section 2.3.5.

since they would not be *about* the concept of interest, grammaticality (Devitt, 2010a, 843).[9] "In sum, the only thing of interest that a speaker could convey to a linguist in [a folk-]acceptability judgment is an opinion about grammaticality" (Devitt, 2010a, 844). This is in line with Devitt's view that the concept one applies in a linguistic intuitive judgement is highly important.

VoC: In his characterisation of VoC, Devitt notes that generative linguists (at least in the more recent literature) often discuss acceptability judgements rather than grammaticality judgements (Devitt, 2006c, 488-490). Based on this, he notes that generative linguists aim to collect acceptability judgements when they elicit linguistic intuitive judgements.

However, Devitt thinks that, since lay speakers are not familiar with the technical notion of acceptability used by linguists, lay speakers are most likely to be giving grammaticality judgements even when asked for acceptability judgements. So while he does recognise generative linguists' intention to collect acceptability judgements, he thinks they do not succeed but are rather collecting grammaticality judgements.

The Critics' Views: On the Critics' Views, the most commonly expressed opinion is that *acceptability* judgements are the kind of intuitive judgements that are relevant as evidence for linguistic theories. On this view, although grammaticality judgements informed by good theories of grammar might make for interesting and helpful hypotheses about language, they are not the kind of data that linguists should aim to account for with their theories, whether those of experts or naive speakers (Culbertson and Gross, 2009; Fitzgerald, 2010; Collins, 2008; Textor, 2009). Not because experts and naive speakers use different concepts of grammaticality, but because we should not use the theoretical "hunches" (Fiengo, 2003) of anyone as evidence. They also deny that speakers (expert or naive) have true grammaticality judgements (in the non-theoretical, mental-grammar sense) in the first place, save in the form of acceptability judgements. This, in their view, leaves us with acceptability judgements, which they take to be pre-theoretical and obtainable from experts and naive speakers alike. Note also that on this view, it does not matter whether lay speakers know the technical meaning of the term "acceptability" as used by linguists or not, since they are not asked to apply that concept to a sentence. On the Critics' Views, speakers are asked to report their reaction of how natural a sentence sounded, whether they would say it themselves etc., without applying a specific folk or technical concept to that experience.

[9] He also emphasises that in this debate, he is only interested in intuitive judgements of this specific type and not, e.g., intuitive judgements about the ease of processing of a sentence (Devitt 2010a, 843; 2013).

Devitt argues that even though generative linguists are aiming for a pre-theoretical or quasi-atheoretical concept which informants can be instructed to use on the spot (such as "sounds natural"), informants are likely to interpret this to coincide with their (folk) concept of grammaticality. And if they do so, they are not delivering what linguists are aiming to elicit. Generative linguists might counter that this is a practical problem that they can (more or less successfully) attempt to solve, for instance by phrasing instructions carefully or by developing better test materials. Devitt, on the other hand, thinks that what we should be after, in any case, is speakers' intuitive judgements of grammaticality.

2.3.3 The role of competence: supplying content or data?

While on all three views, competence plays some role in the etiology of linguistic intuitive judgements, the views differ on the exact role of competence in the process.

The Modest Explanation: On the Modest Explanation, when a speaker is asked to give their linguistic intuitive judgement about a specific sentence, that speaker parses the sentence (like in a comprehension situation) and then "go[es] in for some quick central-processor reflection upon this experience, deploying her concept of grammaticality, acceptability, or whatever from folk linguistics, to form a judgment. The judgment itself is propositional, of course, but the datum for the judgment is not" (Devitt, 2006a, 594). On this view, competence produces the *data* for a meta-linguistic judgement in the form of the experience of processing the sentence, but that is its only contribution to the intuitive judgement.

On this view, Devitt emphasises, "*the data are linguistic expressions* (and the experiences of using them). According to the Modest Explanation, the competence provides only these data to the intuitions. It does not provide any *information about the data*, any linguistic propositions that could be true or false" (Devitt 2010b, 254; original emphasis).

VoC: On VoC, on the other hand, the speaker's linguistic competence provides the informational content for their linguistic intuitive judgements. There are two aspects to this view. First, on this account, the speaker's linguistic competence plays a special role in the formation of an acceptability judgement by providing some form of special input to the judgement. Since the input comes directly from the speaker's competence, it gives privileged access to truths about the speaker's language. Devitt contrasts this with judgements about touch-typing which are more likely based on the subject's experience with the activity and

not based on any special input from a specialised cognitive module.[10]

Secondly, on this view, the speaker's linguistic competence delivers the *informational content* of the judgement without involvement of the central processor. On VoC, linguistic intuitive judgements are not just central processor judgements based on special input from the speaker's competence. Instead, the central processor is not involved at all, and the competence directly provides the informational content of the judgement (say, the proposition "that sentence sounded fine"). On this view, one representation (the rules in the mind) issues in another representation (the content of the judgement) with no input from the central processor. One sign that Devitt has this second requirement for VoC in mind is found in his response to Maynes and Gross (2013). They suggest that, rather than the speaker's competence delivering the content of the judgement, the content of the judgement might be created by the speaker's central processor reacting to the arrival of a parse or, alternatively, a failure to parse the sentence. That would still allow for a fairly direct connection between the speaker's competence and the content of their intuitive judgements, but the competence would not directly provide the content of the intuitive judgement. In this case, Devitt writes, "*the presence or absence* of [a successful parse] would be the *data* for the central processor's response. So, not VoC again" (Devitt 2013, original emphasis). On VoC, "competence does provide [linguistic propositions that could be true or false]" (Devitt, 2010a, 254).

Furthermore, on VoC it is assumed that the way the speaker's competence provides the informational content of intuitive judgements is that the verdict of the judgement is deduced from rules which are somehow implemented in the speaker's mind (more on this in section 2.3.6 below). This is the only specific proposal that I know of for how the speaker's competence could supply the informational content of intuitive judgements (rather than the data for such judgements).

The Critics' Views: Some of the critics argue for a middle way. They propose that although the content of the intuitive judgement is shaped by the speaker's linguistic competence, the contribution of the speaker's competence may not be in the same form as the informational content of the intuitive judgement itself. Instead, they suggest that the output of the speaker's linguistic competence may somehow be fairly directly translatable into the form of a propositional linguistic intuitive judgement. Maynes and Gross (2013), building on a proposal by Rey (2013), outline a model (which is further developed in

[10] Of course, one might hold the view that judgements about acceptability are, in fact, very much like judgements about touch-typing, based on our experience with the activity. As we just saw, Devitt himself argues for such a view.

Gross forthcoming) of how this might work: When a speaker is asked to give a linguistic intuitive judgement about a specific sentence, the speaker's competence is involved in parsing the sentence and creating a structural description of that sentence. If a structural description was made, this will result in the verdict "acceptable" and if none was made, that will result in the verdict "unacceptable" (or a feeling or something similar that could be represented in those terms, see Sprouse 2018 for such a proposal outside this debate). Thus, the output of the linguistic competence is not identical to the content of the intuitive judgement, but there is a close connection between the two nonetheless. Similarly, Textor (2009) describes the intuitive judgements linguists use as evidence for their theories as seemings (a sort of cognitive, non-judgemental evaluations) or feelings which are then reported in the form of a linguistic intuitive judgement with some propositional content. On these views, competence provides either a structural description or a seeming, and this is then the main contribution to the final intuitive judgement, with other cognitive systems merely rendering the input from the competence conceptual (Maynes and Gross, 2013, 719). In the words of Miščević (2006, 546): "It is not that the answer bypasses the central processor, it is rather that the central processor adds nothing substantial."

On these views, there is some role for the central processor to play, although it is a somewhat smaller role than on the Modest Explanation.

2.3.4 Direct access?

Another main question in the debate between Devitt and his critics is whether believing that competence relatively directly gives rise to linguistic intuitive judgements entails the belief that our access to facts about language through intuitive judgements is what Devitt (2006c) calls "Cartesian", meaning direct, a priori, and non-empirical. All sides of the debate agree that judgements might still be influenced by performance factors, however, just like our competence can be influenced by performance factors when we are producing or comprehending utterances (Devitt, 2006c). For this reason, no one thinks that speakers in practice have infallible access to these facts, but if we imagine that performance factors could be filtered out then, on some views, that access would hypothetically be infallible.

The Modest Explanation: On Devitt's Modest Explanation, there is no direct, Cartesian access to information about language through speakers' linguistic intuitive judgements. As described above, the informational content of linguistic intuitive judgements is reached using an empirical theory and central processor reflection, and the speaker's competence only supplies the data on

which the intuitive judgement is based.

VoC: On VoC, Devitt argues, the fact that competence supplies the informational content of linguistic intuitive judgements means that through these intuitive judgements, we have direct, Cartesian access to facts about language. On VoC, the propositions expressed as linguistic intuitive judgements, will, noise aside, be true because they are derived from the speaker's linguistic competence.

The Critics' Views: Few critics address head-on the question of whether they believe that we as native speakers have direct access to facts about the languages we speak through our linguistic intuitive judgements. Miščević (2006) is one exception. He claims that the view he is arguing for does not have the consequence that speakers have direct, Cartesian access to facts about their language through intuitive judgements. Rey (2013) is another exception. In contrast to Miščević, he embraces the predicate "Cartesian" but emphasises that his view is still naturalist in that our access to "what a sentence sounds like to us" is comparable to our access to what some visual stimulus looks like to us (see Devitt 2006c, Devitt 2013, Rey 2013, and Rey forthcoming for a discussion of the analogy to vision).

I think there are signs in the literature that at least some of the other critics, like Miščević, also reject the idea that intuitive judgements give us direct access to truths about speakers' languages. One sign is that their views on what exactly constitutes the data provided by intuitive judgements seem to differ from Devitt's. On VoC (as well as on the Modest Explanation), the *content* of the intuitive judgement is taken to be (likely to be) true *about the speaker's language*. On at least some of the critics' views, on the other hand, the fact that the speaker has a specific intuitive judgement is taken to be evidence *about the system that produced it* (and so not necessarily directly true of the abstract language system itself). Collins (2008, 16) gives this example of a set of sentences:

(3) a. The horse raced past the barn fell.
    b. The paint daubed on the wall stank.

When speakers (who have not seen these examples before) read sentence 3a, their immediate reaction is usually to say that it seems weird. They do not have a problem with sentence 3b, though, even though the two sentences are arguably structurally the same. Once one has identified the correct parsing of sentence 3a, there is in fact nothing wrong with it. Collins (2008) argues that speakers' acceptability judgements about 3a can tell us something about the system that produced the judgement even though the content of the judgement

("that sentence seems weird") is "wrong" about the sentence's grammatical status (although in most normal cases, we would expect the content of intuitive judgements to be roughly "right" about the language system). I take Fitzgerald (2010, 144) to be saying something similar when he states that "[w]e have to make a theoretical inference from a speaker's judgements of acceptability and interpretability to the structure of the underlying competence and its place within wider performance systems." The content of the intuitive judgements delivered by speakers is not necessarily expected to be *true* when taken to be about the language in question, they are rather taken to be behavioural responses to stimuli from which we can tell something about the system that caused the response (Schütze 2006 presents a similar view in a different debate).

On this view, it is not the informational content of an intuitive judgement that is the evidence that linguists rely on, rather it is the fact that the speaker produced such a judgement.[11]

It is possible that it is this disagreement over whether intuitive judgements need to be *about* grammaticality in order for them to be used as evidence *for* grammaticality that is behind the following point of Devitt's. He asks why it is that if, as on VoC, competence speaks, why does it say so little (Devitt, 2006c, 489), i.e., why does it only speak in terms of (un)acceptability/grammaticality, why does it not also speak in terms of other theoretical constructs like "transitivity, heads, A'-positions, c-command, cases, transformations, and so on"? A proponent of the Critics' Views might say that intuitive judgements do not have to be *about*, say, c-command to be evidence *of* c-command.

Devitt might propose different counters to this. On his view, if these intuitive judgements are not *about* the concept in question, they are not the kind of evidence we are (or should be) discussing. In a different context within this debate, he writes that "[t]o the extent that [the intuitions that linguists elicit] are not intuitions about grammaticality, they are evidentially irrelevant to grammars" (Devitt, 2010a, 844). Another line one could pursue would be to say that the critics are simply wrong, and that the content of the intuitive judgements described on the Critics' Views are still *about* the well-formedness of language (i.e., grammar) and that this, contrary to what the critics say, is what matters for the evidential status of these intuitive judgements even if some intuitive judgements are wrong (like the judgement that "the horse raced past the barn fell" sounds weird).

To sum up, the Critics' Views are not entirely clear on this point, and one

[11]But this difference in opinions is also connected to a more general disagreement about whether or not to adopt a mentalist position on what the subject matter of linguistics is, more on this in section 2.3.5. As we shall see in section 2.3.5 neither Devitt's critics nor what seems to be most of generative linguists are against this kind of "psychologising the evidence" (Williamson, 2007).

reason for this might be that the critics are split on the issue of whether intuitive judgements give us direct access to truths about speakers' languages or not.

2.3.5 Conception of linguistics

Devitt and his critics also disagree over what the aim of grammatical research ultimately is. While this issue is not strictly about the nature and etiology of intuitive judgements, it is related to the views adopted by the different sides of the debate, and so the issue is included here. Devitt adopts a non-mentalist position, whereas the critics adopt a mentalist position. On a non-mentalist conception of grammar, the aim of grammatical research is to account for the external patterns we can observe in languages, whereas on a mentalist conception of grammar, the aim of grammatical research is to account for the internal mental mechanisms that give rise to the externally observable linguistic patterns. Fitzgerald (2010, 124) describes this difference of interests in connection to intuitive judgements as data by describing that when psychologists collect intuitive judgements, it is usually because they are interested in what these intuitive judgements "reveal about the psychological states of the people that have them", while at least some philosophers might instead be interested in intuitive judgements "because they are revelatory of a non-psychological domain of facts". These views correspond to a mentalist and a non-mentalist conception of grammar respectively.

The Modest Explanation: According to the Modest Explanation, we should have a non-mentalist conception of grammar, as on this view, linguistics is about the external linguistic reality we can observe in the form of spoken/signed and written utterances (see, e.g., Devitt 2006c, 483). According to Devitt, this linguistic reality is worth investigating in its own right independently of the underlying mental system. On the Modest Explanation, the language system should, however, still be described in a way that is in accordance with what we know about the mind (he calls this the respect constraint, see, e.g., Devitt 2006b, 25). One result of this non-mentalist approach, according to Devitt, is that linguists should limit the use of intuitive judgements as evidence and instead use more direct evidence for grammatical theories, for instance elicited language use.

The Critics' Views and VoC: The mentalist conception of grammar is central to generative linguistics, mainly as an almost pre-theoretic delineation of its area of inquiry: the psychological/mental machinery that allows humans to learn and use natural languages (see, e.g., Chomsky 1995). Not surprisingly,

proponents of the Critics' Views argue that we should have a mentalist conception of grammar. This makes it the main aim of linguistics to model the psychological states and processes that result in the linguistic patterns that we can observe (see, e.g., Maynes 2012a, 445) in a way that allows us to account for what all languages have in common, whatever that might be, and how languages are acquired. This is also the case on Devitt's characterisation of VoC.

2.3.6 Implemented structure rules

Another disagreement between Devitt and his opponents in the etiology debate is over how the mental linguistic system is organised and, relatedly, over what linguistic intuitive judgements may tell us about its organisation. Again, this issue is not strictly about intuitive judgements, but the views of the different sides of the debate relate to their views on linguistic intuitive judgements.

Devitt (2006b, ch. 3) distinguishes structure rules from processing rules. He points out that the rules governing the structure of the output of a competence (the structure rules), must not necessarily be included among the rules governing the exercise of that competence (the processing rules). He argues that in case that the set of processing rules do not include the structure rules, the processing rules must at least respect the structure rules, that is, yield output that conforms to the structure rules, but there are no restrictions on how the processing rules must work to make it so. There is no detectable difference in the output, and therefore we cannot say from merely identifying the structure rules of some output whether or not these rules are included among the processing rules or whether they are merely respected by the processing rules. In the case of language, structure rules would be the generalisations and rules linguists approximate when studying a language as it is used and understood by its speakers, whereas processing rules would be the generalisations and rules that psychologists, neuroscientists, and psycholinguists approximate when studying the part of the mind/brain that deals with processing language. Structure rules could additionally be implemented in the minds of speakers and influence the processing rules without themselves being among the processing rules (see Devitt 2006b, ch. 3-4, for more on this). One question, then, is whether structure rules are implemented in the minds of speakers or not.

Separately from the question of whether structure rules are implemented in the minds of speakers, one can also ask *how* the processing rules are implemented. One possibility is that the processing rules are represented in the mind of the speaker, another option is that they are merely embodied in the mind of the speaker. Devitt compares this to the difference between a general-purpose computer that has a calculator program installed and a specialised pocket cal-

culator. In the first case, the general-purpose computer reads the represented rules in the calculator program when loading and executing the program. In the second case, the processing rules that govern the execution of calculations are not represented in the architecture of the calculator, they are merely embodied as part of the way the calculator is wired. Devitt adds that there is no special justification for positing represented rules rather than embodied rules except when positing representations somehow adds explanatory value.

The Modest Explanation: On the Modest Explanation, the structure rules that linguists aim to discover about languages by studying data like linguistic intuitive judgements do not necessarily tell us how the part of the mind that is responsible for language is organised. Structure rules, on this view, need not be included among the processing rules or otherwise be implemented in speakers' minds/brains. Instead, they need only be *respected* by the processing rules of the mind (Devitt, 2006c). Whatever those processing rules look like, there must however at some stage be *some* embodied processing rules that are responsible for generating the linguistic data for meta-linguistic judgements (Devitt, 2006c, 506).

VoC: On the standard version of VoC (recall section 2.2.1), structure rules are considered to be implemented in speakers' minds/brains, and they are hypothesised to be represented. One reason that Devitt characterises VoC in this way is that traditional generative literature speaks explicitly of mental representation of rules. Another reason for attributing this view to generative linguists might be that it fits well with another part of VoC, namely the idea that competence provides the informational content for intuitive judgements, as discussed in section 2.3.3. If the structure rules governing the output of our linguistic competence are represented in the mind, then the question of how competence could provide the content of intuitive judgements becomes a question of representations of one sort leading to representations of another sort (Devitt, 2006c, 484), which seems like a relatively straightforward hypothesis.

As mentioned in section 2.2.1, Devitt (2006b, 97; 2010a, 836) also considers a non-standard version of VoC on which structure rules are embodied in the minds of speakers rather than represented. He finds it unlikely, however, that embodied rules could give rise to the informational content of linguistic intuitive judgements without some influence from the central processor.[12] Consider the analogy to the pocket calculator: If the rules are embodied rather than represented, they cannot be "read" by another system (the central processor).

[12]He does not think the proposal of Rey (2013) answers this issue, as on Rey's view the content of the intuitive judgement is not provided by the speaker's competence, see Devitt (2013). Rey's view is presented in more detail in chapter 8.

The question Devitt poses is how they could then give rise to the informational content of judgements.

The Critics' Views: Several of the critics seem to hold the view that structure rules are not necessarily themselves implemented in the minds of speakers. Collins (2008, 22) states that he takes the structure rules as identified by linguistic research to be respected by the processing rules of the language. Maynes and Gross (2013, 717), in discussing the evidential role of intuitive judgements, write that linguistic intuitive judgements "provide one important source of evidence concerning these mechanisms [crucial to linguistic production and comprehension] – in particular, for constraints governing their computational operations". Outside the etiology debate, Sprouse and Almeida (2013b) write that they believe that structure rules are at the computational level of Marr's (1982) three levels of analysis (see also Sprouse 2018). I take these statements to mean that at least some proponents of the Critics' Views (and some generative linguists outside the debate) think that structure rules should merely be seen as respected by the processing rules of the mind and should not be seen as necessarily implemented in the minds of speakers (which would mean that they agree with Devitt, at least on this point).[13] Rey (2013) proposes a different view, however. On his view, structure rule *are* implemented in the minds of speakers.

As for whether the processing rules are represented or embodied, on at least some of the critics' views, the linguistic processing rules of the mind are embodied rather than represented (Miščević, 2006; Fitzgerald, 2010; Rey, 2013). Textor (2009, 396) is an exception. He lays out the view that linguistic intuitive judgements are "derived from mentally represented or tacitly known grammatical principles".

So on both the questions of whether structure rules are implemented in the minds of speakers or not and whether the processing rules that are implemented are represented or embodied, opinions are divided among the critics.

2.4 Summing up

Table 2.1 (repeated below) presents an overview of the disagreements between the three views discussed throughout this chapter. These will form the basis for the questionnaire presented in part II of the dissertation.

As can be seen from the table, VoC and the Critics' Views have many features in common, meaning that Devitt, through VoC, has given a characterisation of the point of view of generative linguists that they will agree on for some points

[13] See Botha (1981) for an overview of the then prevailing weaker and stronger stances on the "psychological reality" of linguistic rules within generative linguistics.

Variables / Views	Modest Explanation	VoC	Critics' Views
Origin: experience or competence	Experience	Competence	Competence
Acceptability or grammaticality	Grammaticality	Acceptability	Acceptability
Content or data from competence	Data	Content	Data
Direct access	No	Yes	-
Conception of linguistics	Non-mentalist	Mentalist	Mentalist
Implemented structure rules	No	Yes	-

Table 2.1 (repeated): The three views

(e.g., that intuitive judgements can serve as evidence because they originate in the speaker's linguistic competence, or that we should have a mentalist conception of grammar). However, at least some generative linguists do resist the claim that VoC as a whole is the received generative view on linguistic intuitive judgements. Here, I briefly restate the main points of VoC that Devitt's critics have disagreed with and their proposed alternatives.

Rather than speakers' competence providing the informational content of intuitive judgements, Devitt's critics think that competence provides data for intuitive judgements. At least some of them also do not think that native speakers have direct, Cartesian access to facts about their languages through their intuitive judgements, although there is disagreement within the group on this question. Likewise, at least some of the critics do no seem to think that structure rules must necessarily be implemented in the minds of speakers (though some critics do think that they are).

In the questionnaire, which is based in part on the issues of this debate (see part II of this dissertation), all points that divide the three views discussed above are included, but it will be of special interest to investigate where generative linguists fall on the areas of disagreement between VoC and the Critics' Views as summarised here.

In this chapter, we have seen three different etiology stories about how linguistic intuitive judgements might be related to speakers' linguistic competence. These three views all lead to different answers to the justification question of why linguistic intuitive judgements can be used as evidence for grammatical theories. On the Modest Explanation, intuitive judgements are everyday, theory-laden judgements based on speakers' experience with their language and their (folk) linguistic theories. On this view, the evidential use of linguistic intuitive judgements is justified because speakers have a good amount of empirical experience with their language and relatively dependable theories to build their judgements on. On VoC, intuitive judgements get their content more or less directly from the speaker's competence (through a deduction-like process). This gives us relatively direct access to truths about our languages through our intuitive judgements, and this is why we are justified in using linguistic intuitive

judgements as evidence in linguistics. On the Critics' Views, the evidential use of linguistic intuitive judgements is justified because the speaker's linguistic competence plays a central role in the etiology of linguistic intuitive judgements by providing special input for the intuitive judgements, but the competence does not provide the informational content of judgements.

Chapter 3

The Experimental Syntax debate

The last chapter presented an overview of a discussion of why linguists are justified in using intuitive judgements as evidence for grammatical theories. This chapter looks into the discussion within generative grammar of *how* intuitive judgements should be collected and analysed in order for them to serve as evidence in this way.

Linguistic intuitive judgements are probably the most widely used source of data within generative syntax, and the field is witnessing an ongoing debate on the methodology of how these intuitive judgements are collected and analysed. For a long time, intuitive judgements have generally been collected informally from just one or a few native speakers. In some cases, when the linguist was working on their own native language, they used their own intuitive judgements along with those of their colleagues as evidence. In the debate outlined in this chapter, one side is defending this traditional practice. The other side, the Experimental Syntax (XSyn) movement, argues for a methodological reform. The experimentalists believe that intuitive judgements should be considered a type of experiment and that they should be collected in carefully designed studies, preferably using large numbers of naive speakers and large samples of test items.

In this chapter, I start out by giving a short characterisation of the traditional method as well as of the experimental methods proposed by the Experimental Syntax movement.[1] I then summarise the arguments presented by the experimentalists in favour of using more experimental methods, and then I turn to

[1] This chapter focuses exclusively on the methodological issues surrounding the use of intuitive judgements as evidence for grammatical theories in linguistics. Other linguistic methods and their methodological advantages and disadvantages will not be discussed here.

the counter-arguments presented by the traditionalists. Finally I review some responses from experimentalists to the traditionalists' arguments. In the last section of the chapter, I discuss this debate's relevance for the etiology debate, introduced in chapter 2.

3.1 The traditional method

The way linguistic intuitive judgements have traditionally been collected in linguistics can, in short, be characterised like this: A linguist constructs a sentence that contains some syntactic phenomenon that they are interested in (and which they hypothesise to be grammatical or ungrammatical).[2] They then ask a native speaker, most commonly themselves, whether the sentence appears acceptable or not. Often, they might ask for the opinion of one or more colleagues as well and refine their analysis based on the colleagues' responses (Cowart 1997 calls this the "Hey, Sally" method). The sentence might be considered in the context of another similar sentence which differ from the focus sentence mainly regarding the phenomenon of interest (with the two sentences being called a minimal or near-minimal pair). This is done to increase the likelihood that the judgement of (un)acceptability is actually due to the phenomenon under investigation and not to some irrelevant factor like word frequency, lexical effects etc. Linguistic intuitive judgements obtained this way are often referred to as the data type of choice in generative syntax. Featherston (2007, 275) references Newmeyer (1983) for saying that what a linguists does is "trusting his own intuitions more than any other source of data."

Schütze and Sprouse (2013, 30) mention five ways in which traditional judgement "experiments" are different from standard practice in the neighbouring field of experimental psychology:

i. "relatively few speakers (fewer than ten),

ii. linguists themselves as the participants,

iii. relatively impoverished response options (such as just 'acceptable', 'unacceptable', and perhaps 'marginal'),

iv. relatively few tokens of the structure of interest, and

v. relatively unsystematic data analysis."

It is this informal method of collecting and analysing intuitive judgements that the experimentalists argue against.

[2]The distinction between grammaticality and acceptability will not be defined here, but see section 2.3.2 or, e.g., Gross and Culbertson (2011).

3.2 Experimental Syntax

Schütze (2006/1996) presents an early, comprehensive discussion of the use of linguistic intuitive judgements as evidence for grammatical theories as well as of the problems with the traditional method. Cowart (1997), another early work in the XSyn literature, focuses on how you can design and run syntactic experiments to avoid the problems of the traditional method. The move away from the informal method of data collection within syntax is called Experimental Syntax after Cowart's book (1997).

According to Cowart (1997, 13), the basic experimental set-up for collecting linguistic intuitive judgements should consist of a questionnaire and use multiple informants and multiple test sentences for each phenomenon being studied. To avoid that results are based on irrelevant factors (e.g., subjects judging a sentence to be unacceptable because it contains an infrequent word that they might not know or because they find the situation described by the sentence implausible), different test sentences that relate to the same phenomenon should come in paradigm-like token sets. In these sets, sentences should be as similar to each other as the experiment design allows except for the phenomenon of interest (as with the near minimal pairs described for the traditional method above). That way, if one sentence in the set is judged to be better than the other ones, it will most likely be due to the phenomenon of interest and not some other, irrelevant effect, since the sentences are matched on all other aspects. Furthermore, there should be multiple such sets of sentences for each phenomenon investigated. Cowart further proposes that each test sentence should be rated on a scale but the exact type of scale used might depend on the particular experiment. He also proposes that questionnaires should be constructed based on a counterbalancing scheme to ensure that each participant judges exactly one member of each token set and judges representatives of all sentence types. Each member of each token set should be judged by some subset of participants. Furthermore, Cowart (1997) recommends using filler sentences mixed in with the target sentences and varying the order of presentation between participants when possible. Finally, the results should be subjected to relevant statistical tests.

Similar recommendations for experimental work in syntax are found in Schütze (2006/1996), Wasow and Arnold (2005), Featherston (2007), and Gibson et al. (2013). Another recommendation which reoccurs in the XSyn literature is that subjects should not know what the hypothesis being tested is, and so linguists should not use their own intuitive judgements as evidence (see, e.g., Wasow and Arnold 2005). The benefits and drawbacks of different judgement tasks (forced choice, Likert-scale judgements, the magnitude estimation task, etc.) are described in Schütze and Sprouse (2013) and Sprouse (2018).

Exactly what one should ask participants when eliciting intuitive judgements from them is less extensively discussed in the XSyn literature. Featherston (2007, 292) is one exception. He suggests that the task should "avoid reference to informants' own production" to reduce the risk that participants base their answers on normative factors. As a suitable prompt for eliciting intuitive judgements, he proposes: "How natural does this sound?". According to him, this phrasing has the benefits that it does not relate to frequency of occurrence, and it directs the participant to focus on the spoken rather than the written form of the sentence.

While the instructions are often not discussed in detail, one can find different examples of instructions given to participants in the XSyn literature. Wasow and Arnold (2005, 1491), for instance, report using the following prompt to elicit intuitive judgements from participants: "Rely on your own intuitions of what sounds good, not on what you think is correct according to the experts". Gibson and Fedorenko (2013, 113) describe instructions they used in one experiment in the following way: "the participant was asked to provide a rating for the sentence in the context [...]. There were five choices for each sentence: 'Extremely unnatural', 'Somewhat unnatural', 'Possible', 'Somewhat natural', and 'Extremely natural'."

Perhaps Cowart (1997, 56) is representative of the general attitude within the XSyn camp when he writes that the use of rather informal instructions to participants does not reflect "a lack of concern for achieving uncontaminated results". Instead, he writes, "the emphasis is on experimental control rather than direct manipulation of the informant's approach to the task". In other words, rather than explicitly directing participants' attention to specific aspects of the sentence that they should or should not respond to, test materials should be constructed in such a way that the aspect of the sentence that participants naturally respond to is the one that the investigator is interested in (this is, of course, easier said than done, as presumably requires some amount of testing of the materials before the actual experiment is conducted).

3.3 Motivations behind the experimentalist movement

The experimentalists argue that a change from the traditional method to an experimental approach as described above will ensure studies that are more sensitive and give a more nuanced picture of acceptability, for instance by making it possible to study different levels of acceptability (gradience) and to capture weak effects (see, for instance, Ferreira 2005; Featherston 2007; Haider 2009;

Gibson et al. 2013).

Proponents of XSyn also argue that adopting experimental methods will lead to more reliable data, and thus greater generalisability, by weeding out error variance (random fluctuations across participants). They also argue that it will lead to a higher degree of validity by avoiding, e.g., experimenter bias (see, e.g., Wasow and Arnold 2005; Myers 2009b; Gibson and Fedorenko 2010).

Thirdly, it is argued that a change from the traditional method to experimental methods will yield better methodology and data practices in general. The proponents of XSyn generally criticise proponents of the traditional method for failing to live up to common scientific standards. They argue that achieving better methodological practices would include living up to standards of objectivity (Cowart, 1997; Featherston, 2007; Gibson et al., 2013) and methodological rigour (Wasow and Arnold, 2005; Haider, 2007; Gibson et al., 2011; Schütze, 2011), as well as achieving greater transparency (Cowart, 1997; Myers, 2009b; Gibson et al., 2011).

I go through each of these motivations in greater detail below.[3]

3.3.1 Sensitivity

Schütze (2006/1996) describes how, in the early days of transformational generative grammar, Chomsky suggested that theories should be built on cases where the intuitive judgements of native speakers gave clear results (either clearly acceptable or clearly unacceptable). The theories built on these clear cases would then be allowed to decide the status of the more marginal cases (Schütze, 2006/1996, 23). In Schütze's opinion, the field of syntax has since then moved beyond the clear cases, and theories are now being built on more controversial judgements (see also Ferreira 2005). Therefore, he thinks, the field is now in need of more refined methods which can yield more nuanced data. Haider (2009) agrees and adds that today, competing theories account for the clear cases equally well, and when their predictions differ, the conflict cannot be settled using clear cases. Featherston (2007) agrees with these points and worries that the use of the traditional method for collecting intuitive judgements has led to a situation today where the field of syntax is not making progress (Featherston, 2007, 309-310). For him, "[t]he excitement and sense of progress in grammar nowadays is to be found in work making use of the new qualities and quantities of data

[3]Another point found in the XSyn literature is that syntax should rely less exclusively on intuitive judgement and incorporate more types of data, such as reaction times, monitoring of eye movements, computational and mathematical modelling, or neural imaging. This, it is argued, would allow to calibrate linguistic intuitive judgements. For arguments along these lines, see e.g. Maynes (2012a), Ferreira (2005), and Haider (2009). Since this point does not directly bear on the current debate on *how* to collect and interpret intuitive judgements, it is not discussed further here.

which have become available" (Featherston, 2007, 27). The point advanced by the experimentalists is thus that the traditional method is not sensitive enough to deal with difficult cases.

Another complaint against the sensitivity of the traditional method is that some phenomena may go unnoticed or be underestimated in traditional syntax (Cowart, 1997, 27) and that XSyn is better equipped to bring these phenomena to light. Featherston (2007, 275) puts forward the point that the traditional method does not cast sufficient light on cases where multiple effects are at work at the same time and interact with each other or in cases where the effects that we are interested in studying are relatively small (see also Myers 2009b).

Gibson et al. (2013, 6-7) point out that experimental methods also more easily allow comparing multiple sentences to each other. They argue that even if all the intuitive judgements in a paper containing a small sets of sentences are correct, the traditional method does not easily allow one to compare between the sets. It might turn out that the better sentence in one set is still far worse than the worse sentence in a different set. Also, by using consistent methods, results from different research projects can more easily be compared to each other (Schütze, 2006/1996).

Sprouse (2015, 94) adds that, apart from providing evidence about linguistic structures, XSyn is also well-equipped to study the reliability of the traditional method to find out whether and how the traditional practice needs to change (as we will see below, Sprouse has himself used experimental methods to study exactly this question). He argues that experimental methods can also be used to test reductionist hypotheses about the sources of linguistic intuitive judgements (hypotheses that attribute some effect in intuitive judgement studies to non-grammatical factors, such as working memory) against syntactic hypotheses (hypotheses that attribute effects to the speaker's grammar).

3.3.1.1 Gradience

Another point that Featherston (2007) makes is that experimental methods are better suited to capture subtle, gradient differences in linguistic intuitive judgements. Such gradience could be interesting for two reasons. On the one hand, it might be that the underlying mental grammar is categorical, making sentences either decidedly grammatical or decidedly ungrammatical. In that case, the gradience in linguistic intuitive judgements would come from other, interacting mental systems, and studying the gradience of judgements could teach us something about how the grammar and those systems interact. On the other hand, it might be the case that the underlying grammar itself is graded in nature. Then, of course, studying the gradience of intuitive judgements would

teach us more directly about the grammar itself (see Sprouse 2018 for a thorough discussion of this topic).

Featherston (2007) takes the second approach. According to him, rather than starting out by assuming that the grammar is categorical and assuming that all gradience effects are due to other mental systems, we should take as our starting hypothesis that intuitive judgements reflect the speakers grammar and only subtract certain effects when we have good empirical reason to suspect that those effects are caused by other mental systems. So, to Featherston, experimental methods have the advantage over the traditional method that they give us insight into the (potentially) graded nature of grammar.

The idea that grammar is graded is controversial though, as we will see in section 3.4.1.1.

3.3.2 Reliability and validity

Reliability refers to the consistency of a measure (consistency across time, across participants, etc.). One speaker's linguistic intuitive judgements would be a reliable measure of that speaker's mental grammar if they could be shown to be consistent (across time, across different lexicalisations for the same linguistic structure, etc.). Validity, on the other hand, refers to how well a measure represents the phenomenon it is intended to capture. A speaker's linguistic intuitive judgements would be a valid measure of their mental grammar if the intuitive judgements could be shown to actually give an accurate representation of that speaker's mental grammar.[4]

One of the main complaints raised against specific studies that are done using the traditional method is that certain judgements reported in those studies are hard or impossible for other native speakers to replicate (i.e., different speakers have conflicting intuitive judgements on certain sentences). This is a problem of reliability, since we would expect (correct) judgements to be replicable across speakers. Among others, Wasow and Arnold (2005) and Featherston (2007) report on specific findings from the traditional syntax literature that they have not been able to replicate using experimental methods.

A number of authors raise concerns about factors that can potentially confound results obtained by the traditional method (Cowart, 1997; Schütze, 2006/1996; Featherston, 2007; Wasow and Arnold, 2005; Gibson and Fedorenko, 2010; Ferreira, 2005). Schütze (2006/1996), for instance, presents a thorough list of subject related factors (e.g., cognitive biases, handedness) as well as task re-

[4]Of course, these two aspects of data quality are related. If a measure is generally unreliable, its validity would be low since it would not capture the phenomenon of interest very well. However, a measure can be highly reliable but not very valid (high reliability is necessary, but not sufficient, for high validity).

lated factors (e.g., instructions, ordering) which may threaten the validity and reliability of the data.

According to Cowart (1997, 31-32), an important factor contributing to the lack of reliability of the traditional method is the fact that this method does not take error variance into account. Error variance is random, irrelevant variation in the data. As an example, he reports a study in which participants were asked to judge five different versions of four sentence types, one of which included a that-trace violation (Cowart, 1997, 33). He then compared the average range of variation within sentence types to the variation between the sentences with that-trace violations on the one hand and the average of the other three sentence types on the other hand. For 90% of the participants, the average range of variation *within* sentence types was larger than the variation *between* the that-trace violation sentences and the average of the other three sentence types without that-trace violations. In other words, for the majority of participants, the within sentence type variation was, on average, larger than an effect which is often the object of study in traditional generative literature. Cowart (1997) argues that traditional studies ignore the potential impact on their data from this irrelevant variation. Instead, they take a single person's single judgement to be a reliable report of the acceptability of that sentence type in that person's language. This, in his view, gives experimental methods an advantage over the traditional method as experimental methods provide better control of the reliability of the data.

Myers (2009b) points out that the traditional method actually has some controls for validity and reliability available. Traditional studies, he writes, are fully compatible with using sets of test sentences that are matched on all aspects except the phenomenon of interest (as recommended by Cowart 1997). Another option available to traditional studies is to include irrelevant factors in the design itself to rule out known potential noise factors (e.g., by presenting a test sentence in different pragmatic contexts to highlight different structural readings). He also notes that, while these tools are in principle available to linguists using the traditional method, they are not necessarily applied consistently. There is a third way to attempt to rule out irrelevant effects in studies, which is to use large samples of speakers and test materials. This method, Myers (2009b, 414-415) argues, is unavailable to the traditional methodology. Using large samples can reduce the impact of confounding factors, thus improving reliability and potentially validity as well. It even has the advantage that it may work on irrelevant factors that the experimenter has not considered explicitly, as idiosyncratic biases or lexical effects may be averaged out if enough participants and/or test sentences are included in the study.

Featherston (2007, 282) writes that, in a way, one may see even traditional

introspection as a type of experiment, but he calls it "an extremely *poorly controlled* experiment" (original emphasis).[5]

Summing up, according to proponents of XSyn, experimental methods are an improvement over the traditional method as they allow for better control of irrelevant factors that affect both the validity and reliability of the data. By controlling these factors, the field can gain in accuracy, and resulting theories can be used more confidently as the foundation for new research.

3.3.3 Methodological standards

Several proponents of XSyn have criticised studies using the traditional method for not adhering to the methodological standards that are common in psychology, cognitive science, and other neighbouring disciplines. For instance, Featherston (2007, 271) advocates living up to normal scientific and academic standards as a goal in itself and accuses practitioners using the traditional method for failing to do so, for instance by relying on their own unconfirmed intuitive judgements as evidence. In his view, adhering to traditional practices is "undermining the reputation of syntax and the respect that the field of syntax should enjoy in the wider academic sphere".

A number of other papers arguing for XSyn echo these concerns. Like Featherston, they emphasise the ties to other disciplines and the status of syntax as a scientific discipline. Wasow and Arnold (2005, 1483-4) argue that syntax should live up to the same criteria for good scientific practice as other disciplines in cognitive science. Gibson and Fedorenko (2010) worry about the scientific standards of the field and argue that, even if cases of bad data are rare, their occurrence will diminish the respect for linguistics in other fields with higher methodological standards. Ferreira (2005, 387) argues that linguistics and the field of syntax should stay closely tied to cognitive science as, in her words, "[t]he study of language should be at the core of human cognitive science". This, to her, includes living up to the methodological standards of neighbouring disciplines. Haider (2009) also thinks that the widespread use of the traditional method is at odds with linguists' view of linguistics as part of the cognitive sciences.

This critique of traditional empirical practices represents what some have called a methodological or experimental turn in generative syntax (Machery and Stich 2013; for a description of a similar development in phonology see Myers 2009a). This is in contrast to Chomsky's early recommendation to focus on theoretical advances rather than on methodology and objectivity. He argued that, at least at the time, the data collected with the traditional method was of

[5] See also Chomsky (1986, 36): "Each such judgement is, in fact, the result of an experiment, one that is poorly designed but rich in the evidence it provides."

a sufficiently high quality for the field to be able to make progress and that "a good case can be made for the view that the natural sciences have, by and large, sought objectivity primarily insofar as it is a tool for gaining insight" (Chomsky, 1965, 19-21). It seems that at least some generative syntacticians believe that we have now come to a point in time where increased quality of data *should* be a priority.

When writing about scientific "standards", the proponents of XSyn seem to have a range of things in mind. The following subsections list a number of worries expressed by the experimentalists under the heading of "scientific standards".

3.3.3.1 Quantitative methods / rigour

In many cases, the cry for higher scientific standards seems to be about using the kind of quantitative methods of data collection and analysis that are used in related disciplines. Featherston (2007, 309) mentions that to live up to normal scientific standards, multiple subjects and multiple items should be used, and test items should be presented in an appropriate context with an appropriate range of answers allowed. Gibson et al. (2011, 509-510) advance a similar list, adding that participants should be blind to the research question and hypothesis, that distractor elements should be included, and that the order of presentation should be considered carefully and vary among participants. They hope that actions like these "will further the adoption of quantitative methods, advancing the empirical standards of linguistics to those found in other psychological sciences" (Gibson et al., 2011, 522). Gibson and Fedorenko (2010, 233-234) and Myers (2009b, 406) present similar concerns.

Some of the proponents of XSyn also criticise the traditional method for not being rigorous enough. In such passages, *rigour* is sometimes equated with the use of quantitative methods (e.g., Gibson et al. 2011, 2013; Schütze 2011) as seen in this quote:

> Much of cognitive science has its roots in the works of philosophers who did not have the tools to evaluate their hypotheses quantitatively and were limited to introspection and "thought experiments". However, in many cases (and certainly in the case of syntactic intuitions), we no longer lack such tools, and thus nothing should stop us from evaluating claims about the human mind with scientific rigor. (Gibson et al., 2013, 2)

Sometimes, *rigour* is also equated with conducting data collection and analysis in a systematic manner, as when Wasow and Arnold (2005, 1486) write

that "unsystematic use of primary intuitions is so entrenched among generative grammarians that contradictory evidence from other sources – including more rigorously collected intuition data – is simply ignored".

Furthermore, the concept of *rigour* is invoked when proponents of XSyn want to advocate using methods that by default provide a certain detachment between researcher and results, as in this quote:

> It is important that syntacticians should show scientific detachment and be seen to be producing theory to fit the facts, rather than the opposite. This cannot be achieved by using our own unconfirmed intuitions. A study which uses data from multiple informants is thus more intellectually rigorous than one which does not. (Featherston, 2007, 283)

See Haider (2007, 389) for a similar comment.

In sum, *rigour* seems to be taken by proponents of XSyn to mean using quantitative methods, working systematically, and being careful about separating hypotheses and findings. All this, they argue, is often lacking from studies using the traditional method whereas it figures more centrally, they believe, in studies using experimental methods.

3.3.3.2 Objectivity

Experimentalists also claim that the traditional method lacks objectivity. For instance, both Cowart (1997) and Featherston (2007) note that objectivity in science boils down to inter-subjectivity, which is achievable in the case of linguistic intuitive judgements by averaging over the judgements of multiple speakers. This objectivity (i.e., inter-subjectivity), they argue, is lacking in studies based on data collected with the traditional method from one or a few speakers. According to Featherston (2007, 278), intuitive judgements collected the traditional way are unconfirmed, and using them therefore constitutes insufficient research practice (even if it were not for other problems with the method). Gibson et al. (2013, 10) put it like this: "Science, in short, seeks *objectivity*, a value that has a rich history in the development of scientific method [sic]" (original emphasis).

3.3.3.3 Transparency

Cowart (1997) mentions that experimental methods include a requirement to account for how the data was collected and analysed, which adds transparency to experimental studies. This is an advantage over the traditional method, he thinks, where there is no similar tradition, and the reader has to trust that the author performed all the relevant checks and controls (Cowart, 1997, 10).

Similarly, Featherston (2007, 282) writes that "[it] is standard practice in the academic world for sources to be identified; there is no reason for introspective judgements to be made an exception to this". Ferreira (2005, 372) advances a similar point when she writes that even if, in a traditional study, the linguist asks a number of colleagues for their intuitive judgements, there is no policy for "reporting and reconciling contradictory opinions".

Gibson et al. (2011) argue that even if the majority of judgements in traditional syntax are correct, the presence of some that are not may damage the field because we do not have a way of knowing which ones are the correct ones. Myers (2009b, 409) observes that when using experimental methods to analyse intuitive judgements, one gets a measure of statistical significance, which, "in turn, is related to the probability of future replications". In this way, experimental methods may add transparency about not just the data collection procedure but also about the data itself. The lack of similar transparency for the traditional method means that a researcher reading about results from a language that they do not themselves speak have no way way of assessing the trustworthiness of the result. Compared to informal judgements, he writes, "[a] formal judgment experiment, with an explicit, replicable methodology, has a better chance of resolving such data disputes" (Myers, 2009b, 409).

3.3.3.4 Data first

Featherston (2007, 315-316) thinks that practitioners of the traditional method are more concerned with coming up with linguistic theories that live up to the theoretical virtues of being "elegant" or "economic" than with coming up with theories that accurately account for the data at hand. He is strongly critical of this perceived trend. He further writes that, within traditional syntax, "the details of gathering and analyzing data are regarded as uninteresting, almost taboo" (Featherston, 2007, 278).

Another worry put forward by the experimentalists is that the traditional methodology does not provide sufficiently strict divides between hypotheses and findings:

> The experiments are necessary in *all* cases because they provide the only way to objectively measure discrepancies between theory and reality. Theories evaluated only by the intuitions of the investigators involved, are almost necessarily *post hoc*. This is because, lacking quantitative standards, we have little possibility to be wrong and discover discrepancies between theory and data. (Gibson et al. 2013, 10, original emphasis)

Featherston seems to be thinking of something along the same lines when he

writes that one benefit of using experimental methods is that "[f]irst and foremost it allows the linguist to avoid the stick of criticism for producing theory without an appropriate basis in data. There can be little satisfaction in producing or reading work which so clearly fails to satisfy scientific and academic standards" (Featherston, 2007, 308).

Finally, Gibson and Fedorenko (2013) recognise that adopting an experimental approach will not in and of itself make confounding factors go away or ensure the correct interpretation of data. They believe, however, that, all other things being equal, a change of methods will minimise the reliance on spurious data.

3.3.4 Summing up

The experimentalists believe that the traditional method of collecting and analysing intuitive judgements is not sensitive enough to capture all relevant effects, that it is influenced by irrelevant factors, sometimes resulting in unreliable and invalid data, and that it fails to live up to common scientific standards.

Compared to using informally collected intuitive judgements, the experimentalists argue that using experimentally collected ones will help ensure:

1. more sensitive studies and thus a more nuanced picture of acceptability, for instance by making it possible to study gradience and to capture weak effects,

2. more reliable data, and thus greater generalisability, by reducing error variance,

3. more validity, e.g., by avoiding cognitive biases, and

4. better methodology and data practices in general, which includes living up to standards of objectivity, methodological rigour, and transparency.

3.4 Traditionalist reactions

Here, I present the reactions of some of the proponents of the traditional method to the experimentalists' criticism. Most of the traditionalist reactions include statements to the effect that the authors do agree to some degree with the sentiment behind the XSyn movement, e.g., that some amount of controls are important for the quality of one's data. So rather than an outright rejection of the experimentalists' position, the traditionalists' responses to the experimentalists express modifications, different interpretations, different priorities etc.

Also, in some cases we find authors agreeing with some of the experimentalist arguments and agreeing with some of the traditionalist arguments as well. These authors figure in both sections of this chapter.

3.4.1 Sensitivity

A few replies to the experimentalist works point out that variance in judgements might provide valuable information that we miss if we only study the average of a large sample of judgements. This is the idiolect-argument against experimental methods, and it is built on the theoretical assumption that speakers of the same language (English, Finnish etc.) may speak different micro-variants or idiolects. On this view, when you average across speakers, you dismiss potentially meaningful micro-variation as noise. In the opinion of the authors in favour of the idiolect-argument, it is a plausible possibility that speakers of, say, some dialect of English have all acquired slightly different grammars, speak their language or dialect in slightly different ways, and that they will therefore also judge the acceptability of sentences in different ways. According to Fanselow (2007, 360), "we cannot attribute such extreme judgment differences [between speakers] to *random* noise effects. They represent stable response patterns on the individual level" (original emphasis). Den Dikken et al. (2007, 341) argue that, contrary to the experimentalists' call for testing many subjects, studying the judgements of a single individual suffices: "UG lives in every individual, and there is in principle nothing wrong with trying to model the grammar of a single individual as a way to study its limits and possibilities". They emphasise that we should only generalise over a group (to weed out potential noise) if we can be confident that they share a grammar. They do acknowledge the existence of noise (Den Dikken et al., 2007, 350-351), but they do not specify if or how this should be dealt with (presumably, it would be detectable if the same speaker is tested multiple times with the same or similar stimuli). But, they argue, one cannot know before looking carefully into a case of variation whether it is noise or the sign of a micro-variant. According to them, variation should not by default be assumed to be noise. They write that "one's judgments are one's judgments, no matter what other speakers of 'the same language' might think" (Den Dikken et al., 2007, 343).

3.4.1.1 Gradience

Some proponents of the traditional method provide counters to the argument that experimental methods better capture the graded nature of grammar through graded linguistic intuitive judgements. Haider (2007) as well as Fanselow (2007) argue that this is not actually a benefit for the experimental methods, since

gradience is more likely to result from extra-grammatical factors like processing difficulty than to be a feature of grammaticality itself. In short, gradience might not be related to the phenomenon we are actually interested in. Haider (2007, 392) argues that even though speakers may be able to use a graded scale to express how different sentences compare to each other, they will still have a categorical answer if asked whether or not a given sentence could be uttered by a native speaker or not (i.e., some sentence might be worse or better than others, but each sentence will always fall into either the category "grammatical" or the category "ungrammatical"). Fanselow (2007) quotes research by Armstrong et al. (1983), finding that even strongly categorical distinctions such as evenness or oddity of numbers may show graded judgement patterns in certain experiments. These judgements in relation to numbers, Fanselow reasons, are influenced by frequency and typicality effects which might also very well be at work in linguistic intuitive judgements.

3.4.2 Reliability and validity

Phillips (2009) questions the factuality of the claim advanced by the experimentalists that the informal collection method and the quality of the resulting data have halted the progress of syntax. Cases of bad (or irreproducible) judgements do exist, he agrees, but he argues that they are easily caught by reviewers and readers. This means, he believes, that they are not allowed to go far in the theoretical debate and thus that they are not a great threat to the development of the field. Therefore it is unlikely, he thinks, that important theoretical work is based on bogus data. He also recounts a number of important, large-scale theoretical questions (e.g., transformational vs. nontransformational grammars, and government-binding theory vs. minimalism) and writes that these disputes are more likely to be based on differences in how to interpret the relevant datasets (including linguistic intuitive judgements) than disputes over what the correct judgements on particular sentence types are.

Sprouse and Almeida (2013a) treat the question of whether the traditional method is unreliable as an empirical question. They investigated well traditional judgements presented in textbooks and journal articles correspond to intuitive judgements elicited in formal experiments (see also Sprouse et al. 2011; Sprouse and Almeida 2012a,b; Sprouse et al. 2013; Sprouse 2015). They argue that the claims that the traditional method is unreliable are themselves made on small, biased samples of examples (basically, any collection of examples of bad judgements that opponents of the traditional method could find). Instead, they think, to truly test those claims, a more rigorous investigation of a broad, representative sample of judgements has to be made. They report on experiments of

exactly this kind that they performed. In these experiments, they take example sentences from a widely-used syntax textbook as well as from a sample of syntax papers from a respected journal and collect intuitive judgements on those sentences from large numbers of ordinary speakers, using the "best practices advocated by [Gibson and Fedorenko (2013)]" (Sprouse and Almeida, 2013a, 224).

They found replication rates of 95-98% and conclude that the problems that are attributed to the traditional method "have not generated an epidemic of unreliable data in syntactic theory" (Sprouse and Almeida, 2013a, 224).[6] They ascribe the high replication rate to the following a) judgements are cheap and easy to replicate and so can be tested informally before publication, and b) phenomena that traditionally interest syntacticians have relatively large effect sizes, i.e., a large difference in acceptability depending on presence or absence of the phenomenon of interest.[7]

Furthermore, they report an experiment that shows that using 10 subjects will give reasonably good statistical power and detect a significant amount of the phenomena reported in the literature.[8] This experiment used a standard method in traditional syntax, the two alternatives forced choice task, in which the participant is presented with two sentences and asked to judge which one is more acceptable (Sprouse and Almeida, 2013a, 225). This practice, it can be argued, comes close to the "Hey, Sally" variety of the traditional method where the investigator asks a number of colleagues. The authors also show that some methods proposed by critics of the traditional approach, for instance magnitude estimation (see, e.g., Bard et al. 1996) and 7-point Likert-scales, have lower statistical power than the two alternatives forced choice method. Taking the two alternatives forced choice method to be representative of a least a substantial part of traditional syntactic methodology, they conclude that experimental methods cannot by default be assumed to be better than the traditional method without a thorough discussion of tasks, statistical power, effect sizes etc.

As for the point that the experimentalists make about the lack of validity of the traditional method, Sprouse and Almeida (2013a) argue that the data they have surveyed suggests that linguist-participants are not overwhelmingly influenced by cognitive bias when making judgements. This they base on the high level of convergence between the judgements of linguists (in the journal papers

[6]Sprouse and Almeida (2013a) note that, as they interpreted all failures to replicate as true negatives and used two-tailed p-values, a replication rate of 95-98% is a conservative estimate. However, as we will see in section 3.5.1 below, Gibson et al. (2013) argue that the replication rate reported by Sprouse and Almeida is, in fact, inflated.

[7]Effect size has an influence on the number of subjects needed to detect an effect. Roughly speaking, if the effect size is larger, fewer subjects are needed, and vice versa.

[8]The power of a test is the probability that it will correctly reject a false null hypothesis when the alternative hypothesis is true.

and textbook examples) and lay subjects (in the experiments). If the linguists were influenced by cognitive bias, Sprouse and Almeida reason, we should see either a high number of replication failures (linguists reporting differences that lay subjects could not detect) or sign reversals (linguists reporting differences between sentences that were judged to be different, but in the opposite direction, by lay speakers). They do not find evidence of either of these issues being frequent in their studies. So while they agree that cognitive bias is a potential problem for the traditional method in principle, they see no evidence of a de facto problem.

Phillips (2009) worries that experimental judgement collection over the internet (using sites such as Amazon Mechanical Turk) might introduce new problems compared to the traditional method, because participants might not receive proper instruction and thus not understand the task they are performing. This would introduce a possible confound which is much less likely to appear in traditional judgement collection, where trained linguists who already understand the task at hand are used as subjects. This would be a problem for many experimental studies, which have recruited their participants in this particular way.

Culicover and Jackendoff's (2010) suggestion is to focus on controlling the stimulus sentences and pay less attention to the other aspects of experimental methods, such as using large numbers of subjects, making statistical analysis etc. Adopting these quantitative standards, they think, would cripple linguistic research because of the dramatic increase in effort needed to perform studies (see Gibson et al. 2011, 510, for a reply to this argument).

3.4.3 Methodological standards

The traditionalists' answer to the experimentalists' call for more rigour and adherence to common methodological standards comes in two overall varieties. One variety agrees in principle that methodological standards are important but disagrees that the traditional method is unscientific. The other variety disagrees altogether that living up to certain methodological standards should be a goal in itself.

On the first side, Newmeyer (2007, 396) agrees "in principle" with some of the points made by Featherston (2007) about applying quantitative standards in syntax but makes the point that more rigorous testing is only necessary if the rewards justify it (either through achieving positive effects or avoiding negative ones). Interestingly, this response accords with the proposals of Featherston (2009). In this paper, compared to Featherston (2007), he takes a more traditional-friendly approach and argues that while methodological rigour on

some points is important, some aspects of experimental settings may not make a great difference in most cases, and therefore these can be ignored in most experiments. He recommends focusing on test materials and worrying less about subjects and specific methodology. Culicover and Jackendoff (2010) likewise argue that while proper control of stimulus material is of major importance in science, using experimental methods is in itself no guarantee that proper data practices are followed (going against the arguments of, e.g., Haider 2007, 2009). Along the same lines, Phillips and Wagers (2007) argue that because the traditional method is so easy and inexpensive to check, not a lot of safeguards on reliability of single data points are needed (contrary to psycholinguistics, where getting trustworthy information about one single fact is laborious, and where more explicit checks and transparency is therefore needed). On this approach, the proponents of XSyn are right in principle, but there is no need to worry in the case of traditional judgement collection as the practice is already satisfying the standards to a reasonable degree today.

The other type of reply is exemplified by Grewendorf (2007), who argues more explicitly against the focus on objectivity and standards of methodology as a goal in itself.

> Instead of taking more care with the data and collecting reliable data, scientists often just disregard the phenomena in order to find principles that seem to give some deep insight into reality, where reality is taken to be represented by the abstract systems that are constructed rather than by an unstructured conglomeration of phenomena. There are numerous examples in the history of the natural sciences, some of them mentioned in Chomsky (2002), which show that criteria such as simplicity, elegance, and fruitfulness are as important in the assessment of theories as are reliable data. (Grewendorf, 2007, 379)

This also goes against the view that accounting for the data should have main priority in syntax (e.g., Featherston 2007).

The experimentalists' calls for more transparency are not explicitly answered by the traditionalists. Sprouse and Almeida (2013a, 227), however, suggest that researchers should be more explicit about a) potential confounds that could have an impact on their results and the interpretation of these, and b) whether there exists empirical evidence that such a confound is actually affecting the results. They thus agree with the general sentiment of the proponents of XSyn regarding transparency.

Some authors downplay the magnitude of the difference between XSyn and traditional syntax. Sprouse (2015), for instance, believes that it is a difference

of degree rather than a difference of kind. He writes that both approaches make use of "tightly controlled experimental conditions" (presumably the presentation of controlled stimulus) to test "the same behavioural response" (intuitive judgements). The difference, he believes, is in the number of participants and stimulus items used (Sprouse, 2015, 89). He does mention one difference of kind though, the background knowledge of participants. In most cases where the traditional method is used, trained linguists are used as subjects, whereas naive subjects tend to be used in most studies done with experimental methods. One worry about the use of experts as subjects is that there might be cognitive biases at work (as mentioned), but Sprouse (2015) points to some of the studies mentioned above that show a very high level of convergence between the results of the traditional end experimental methods. These results suggest that cognitive bias among the linguist-subjects might not have a large effect on the data.

3.4.4 Summing up

The experimentalists' three main objections to the traditional judgement collection method in syntax were related to the lack of sensitivity of the method, the potentially low reliability and validity of the method, and the lack of adherence to common standards in related disciplines (let us call these the sensitivity objection, the reliability objection, the validity objection, and the methodology objection).

The traditionalists' main answer to the sensitivity objection seems to be that while some nuances are fine, it is not at all certain that all details revealed by experimental methods are relevant to the phenomenon of interest, i.e., grammaticality.

The reliability objection is answered by some traditionalists by showing how the traditional method yields highly reproducible results, while others make the point that valuable information on micro-variation may get lost when results are averaged across speakers. The validity objection, when addressed, is answered by claiming that although things like experimenter bias could in theory be problematic, it does not seem to pose a de facto problem for the traditional method.

There seems to be two strategies for answering the methodology objection as well. Some authors think that we can show that the traditional method is fully adequate as it is,[9] and some add that using an experimental procedure is, in and of itself, no guarantee that greater care will be taken with data and theoretical issues. Others instead believe that the main point of a scientific test

[9]Proponents of XSyn might argue, however, that it took experimentalist raising the issue *and* the use of experimental methods to find this out, while this is something users of the traditional method should have worried about themselves all along.

is to give valuable insights, not to be maximally methodologically rigorous or objective according to some standard.

While this discussion has been phrased in terms of proponents and opponents of the traditional method, in reality the discussed works can be seen as aligned along a continuum. Some authors, like Wasow and Arnold (2005), who argue that the traditional method should be supplanted completely with experimental methods, are at the extreme experimental end of the spectrum. Most authors are somewhere in between. Those who are at the more experimental end advocate that experimental methods should be used in general but that in some cases it might be enough to use the traditional method (e.g., Featherston 2009), and those at the more traditional end (like Phillips 2009) argue that the traditional method should in general be used but that in some cases, it might be beneficial to use experimental methods. Interestingly, even the most ardent defenders of the traditional method all seem to agree with at least some parts of the experimentalist programme (at least in principle), and the extreme traditionalist end of the spectrum is thus (at least in principle) unoccupied.

3.5 Experimentalist reactions

3.5.1 Convergence of expert and lay intuitive judgements

Arguably, one of the most interesting results to come out of the discussion between the experimentalists and traditionalists is the work by Sprouse and colleagues showing that there is a high level of convergence between the judgements of lay subjects and those of trained linguists. How do the experimentalists reply to that?

Gibson et al. (2013) take up the discussion with Sprouse and Almeida (2013a) about the reliability of the traditional method. In their opinion, since linguistic theories are meant to get every grammaticality assignment right, it is not necessarily the case that an error rate of 2-5% is acceptable (Gibson et al., 2013, 4). Gibson et al. argue that we want our syntactic theories to account for not just one or a few but a large number of phenomena. However, their calculations show that if we allow for an error rate of 5% (and assume the standard power level of 80%), then we can only build our theory on less than five data points before we are likely to get a theory that is based on incorrect data. If any data point that the theory is built on is wrong, the theory will be wrong in some way or another as well, and non-quantitative methods have no method to detect and correct such errors. Even with low error rates, we have no way of knowing which data points are correct, and there will be incredibly many ways to divide up the set of data, so we will not be able to go through each one. With experimental

methods, on the other hand, one has the option of lowering the false positive rate by adding more participants. Experimental methods not only give us good data, they also tells us something about the reliability of specific sets of data and ways to improve the reliability when necessary.

They also note that the estimates made in Sprouse and Almeida (2013a) based on journal data hinge on a random sampling of every standard acceptability judgement published in the journal Linguistic Inquiry from 2001 to 2010. Thus, they point out, the sample contains a lot of completely uncontroversial judgements (e.g., that "was kissed John" is unacceptable, Gibson et al. 2013, 3) and not just those intuitive judgements that differentiate current theories. For that reason, Gibson et al. think that Sprouse and Almeida's estimate of the convergence between experts and lay subjects is inflated. Therefore, they argue, the sample cannot be used to show that the traditional method can be used to choose between the theories that are debated today, which was one of the original motivations for XSyn.

Gibson et al. (2013) also argue against the argument put forward by Sprouse and Almeida (2013a) that the statistical power needed to detect the type of effects linguists are typically after are usually quite low, and that a small number of judgements may therefore suffice to gain sound results about those effects in general. Gibson et al. (2013) argue that even if most syntactic questions have large effect sizes, one still needs to collect data and calculate the effect size for the specific contrast one is working on to verify that this phenomenon is, in fact, one of the ones with a large effect size. There is no way of knowing this without doing the experimental study. Otherwise, they point out, you get a false sense of confidence in your results (Gibson et al., 2013, 5).

3.5.2 Idiolects

Another interesting discussion to come out of the XSyn debate is the discussion over whether syntacticians should be studying single speakers' idiolects. Wasow and Arnold (2005, 1487) dismiss the idea that syntax should model individuals' idiolects rather than treating variance as possible noise, saying that if judgements are, in principle, not comparable across speakers "then they differ from every other sort of evidence used in science, where replicability of experimental results is normally taken as an essential standard of evaluation".

Featherston (2007, 279) also disagrees with the traditionalists who use the idiolect-argument against experimental methods. He argues that the evidence does not support the hypothesis that people in general have different idiolects:

> The prediction of different grammars would be that we should see separate clusters [of judgements of a given structure] for the infor-

> mants for whom the structure is good and for those for whom it is bad. But this does not occur: all we ever see is variation around a common pattern, plus occasional outliers. (Featherston, 2007, 284)

So, Featherston argues, we do not see the data patterning in the way we would expect it to if variation among speakers was indeed due to systematically varying idiolects (rather than random noise). Likewise, Myers (2009b) claims that there is no evidence that judgement idiolects are real.

Featherston (2007, 284-285) suggests that the idea that idiolects exist may come from the (misguided) practice of comparing two individuals' judgements to each other instead of comparing each judgement to a mean taken over a number of judgements. Since individual judgements are subject to random error variation, comparing two single judgements might give the impression that the subjects disagree over the acceptability of a certain structure when, in reality, the difference might be due to random, irrelevant variation. If one does want to compare the judgements of a single particular speaker against something, it is better to compare the speaker's judgements to the mean values obtained from a group of participants, since in that case the noise is more likely to have cancelled out and error variance would be reduced.

Another possible source of the apparent idiolects, he argues, might be imposing a dichotomy onto material of a graded nature (Featherston, 2007, 285). If two judges both find sentence X to be marginal but are only allowed to answer that it is either "completely acceptable" or "completely unacceptable", a difference in their choice of category might depend more on their (idiosyncratic) choice of categorical answer rather than on them having different intuitive judgements of the sentence.

3.6 Experimental Syntax and the etiology debate

Here, I want to briefly touch on how this discussion may be relevant to the debate about the justification for using intuitive judgements as evidence for grammatical theories in linguistics.

The only work discussed in this chapter that mentions the justification question explicitly is Maynes (2012a). He proposes that linguists can justify the use of intuitive judgements as data by calibrating the results of judgement "experiments" (whether formally or informally conducted) with more linguistic intuitive judgements (to minimise error variance, as already discussed) or with other types of data.[10] This, he argues, allows linguists to rely on intuitive judgements

[10]The calibration approach was mentioned in section 2.1 as one way to approach the justification question along with the "fruitfulness" approach and the etiology approach.

while not yet having an adequate understanding of how the machinery behind the intuitive judgements work. This, in turn, means that linguists may be able to escape Devitt's (2006c) choice between the Voice of Competence view (VoC), his own Modest Explanation (see chapter 2), or acknowledging that we have no good justification for the use of intuitive judgements. To perform calibrations between instances of judgements, we need formal methods. But at least in principle, Maynes' suggestion does not require the entire field of syntax to change to experimental methods since, as shown by, e.g., Sprouse and Almeida (2012a), traditionally collected judgements can also be verified by calibrations (leaving out the question of how to deal with cases where you reach different results through the two methods, Maynes 2012a, 457-458).

Thus, if one approaches the justification debate from Maynes' angle, the choice between experimental methods and the traditional method does not entail a choice between different justifications for the use of intuitive judgements as data.

One might, however, also wonder whether the XSyn debate might be relevant to the etiology debate presented in chapter 2 (which, as mentioned there, is one take on the justification question).

On two points, it might, at least at first glance, seem that traditionalists' views might be related to VoC as characterised by Devitt. 1) Since there is no focus on error variance in traditional syntax, it seems like the field is assuming that asking a single subject yields the "correct" result, which would be the case if intuitive judgements were really the voice of competence. 2) The rejection of the empirical standards of neighbouring disciplines makes it seem like the field views the process of arriving at linguistic intuitive judgements as somehow "special", which means that eliciting them needs not be subject to the same demands of empirical rigour, objectivity, etc. as other scientific tests. This could be due to an assumption of something like what Devitt refers to as direct, Cartesian access.

Looking closer at the views expressed by the traditionalists, however, I do not think that a traditionalist view on syntactic methodology and VoC must necessarily go together. Regarding (1), as we have seen, the traditionalists argue either that the traditional method in fact does deal with error variance (through asking colleagues and through the review of papers), just not as systematically as when using experimental methods, or that variance is potentially meaningful and should be studied rather than abstracted away from. The focus on idiolects is, in and off itself, not more compatible with the view that intuitive judgements give us direct access to the speaker's competence than with the view that intuitive judgements are empirical, metalinguistic judgements based on the speaker's experience with their language.

Similarly with (2), the traditionalists argue that although the methodological standards in traditional syntax are different from those of neighbouring disciplines, this is because the effects of interest in syntax are highly robust ("sledgehammer effects", Schütze 2011), and the current level of methodological rigour is enough to capture the needed results, which makes further focus on methodology a waste of time and energy. Another possible answer, given by some proponents of the traditional method, is that gaining deeper insight, no matter the method, is more important in any field of science than living up to certain standards of objectivity and methodological rigour just for the sake of it. This, again, does not necessarily come with a belief in special, Cartesian access to facts about language through intuitive judgements.

The traditionalists' preference for using the intuitive judgements of linguists over those of lay subjects, in fact, seems to point in the opposite direction. On the Modest Explanation, we should use the intuitive judgements of those with the most expertise and the best theories. However, as it is a general commitment within generative linguistics that all speakers are equally competent with their native language, and as it seems to be a widespread generative assumption that linguistic intuitive judgements express speakers' competence (rather than their linguistic theories), it is unclear whether the traditionalists' motivation for preferring to use experts as subjects is the same as the motivation on the Modest Explanation. In chapter 6, we will learn more about the views of some generative linguists on these topics as we look into the questionnaire data.

As for the experimentalists, it might seem as if there is an obvious connection between experimentalist views on syntactic methodology and the view that intuitive judgements are not the Voice of Competence. However, I do not think that this connection is necessary either. Someone arguing for the use of experimental methods in syntax might believe in VoC but also think that other things, such as performance factors, can affect judgements and that these need to be ruled out by using proper experimental controls.

So overall it does not appear to be the case that any stance on the XSyn debate is necessarily connected to any particular positions in the etiology debate. The fact that there is no conceptually necessary connection between the views in the two debates does not mean, however, that the views might not tend to go together in practice. In chapter 6, we will see if there are any actual patterns of different opinions on the etiology question according to stance on methodology.

Chapter 4

Intuitive judgements in non-generative linguistics

While the evidential use of linguistic intuitive judgements is more widespread within generative linguistics than in other linguistic frameworks, this project also aims to shed light on how the use of linguistic intuitive judgements might be justified within other traditions of linguistics.

In this chapter, I review how intuitive judgements of morphosyntactic well-formedness are used and perceived within the literature of non-generative branches of linguistics. There, it is less extensively discussed than in the generative literature, and we therefore know less about the views of non-generative linguists on this topic than we do about the views of generative linguists. In chapter 6, we see how the non-generative linguists in the quantitative study answer a questionnaire based on the main themes discussed in the two preceding chapters.

4.1 Intuitive judgements as evidence in fieldwork

When studying a little- or previously undescribed language "in the field", linguists try to systematise the utterances they encounter to get an overview of the grammar of the language in question (or, of the subset of the grammar that they are working on). One method that linguists can rely on to try to patch apparent gaps in their corpus of encountered utterances is to construct sentences they think should (or should not) occur in the language and ask one or more native speakers whether this sentence is something that a native speaker of their language might say or not. They will then have to examine the answers carefully to try and find out if, for instance, a negative answer might be due

to lexical choice, register, politeness issues etc., or whether it is best analysed as being due to the grammar of the language. This quick sketch aims to show how linguistic intuitive judgements by native speakers can be of great value to linguistic fieldworkers as a source of evidence for grammatical descriptions.

Several introductory texts on the methodology of linguistic fieldwork mention linguistic intuitive judgements as one source of evidence that the field linguist may attempt to elicit from their native speaker consultants. For instance, Majid (2012, 54) writes: "The linguist might use a word list or questionnaire and ask a consultant, 'How do you say X?', probe for grammaticality judgements, or solicit translations." Chelliah (2013, 60) also mentions linguistic intuitive judgements as one source of evidence for the linguistic fieldworker: "For example, the fieldworker may construct a sentence in the target language and ask for a grammaticality judgement or other introspective statement. This is the classic 'Can you say X?' Or 'Is X a possible sentence in your language?' method of elicitation".

For linguistic fieldworkers, elaborate experimental studies are often not feasible, for instance when the fieldworker only has access to a limited amount of speakers, only has a limited time available with the speakers, if the speakers are unable to read or write (or, in general, if the speakers are not from a WEIRD background: Western, Industrialised, Rich, Democratic), and due to other limits on resources. In the case of languages with only a handful of living speakers, proper formal experiments might even be impossible. This means that the intuitive judgements elicited in linguistic fieldwork are likely to often be closer to the informal end of the scale than to the formal experimental methods described in chapter 3.

This method of elicitation is used across different theoretical orientations and is not limited to field linguists working within the tradition of generative linguistics. It is thus not necessarily tied up to any particular theoretical assumptions about the etiology of intuitive judgements or any specific justification for using them as evidence. Rather, these judgements may be seen as simply a practical, if not bullet-proof, way to gain some information on structures that one is otherwise lacking data on.

4.2 Intuitive judgements in non-generative frameworks

Linguistics is a fundamentally divided discipline, as far as theoretical foundations and empirical methodology are concerned. On the one hand and with some simplification, there is the field of genera-

> tive grammar with its assumptions of (i) a highly modular linguistic system within a highly modular cognitive system (ii) with considerable innate structure given the poverty of the stimulus, and (iii) a methodology largely based on made-up judgments of made-up (often context-free) sentences. On the other hand and with just as much simplification, there is the field of cognitive/functional linguistics with its emphasis on (i) domain-general mechanisms, (ii) pattern-learning based on statistical properties of the input, and (iii) an (increasing) reliance on various sorts of both experimental and observational data. (Gries, 2014, 15-16)

This characterisation by Stefan Th. Gries seems to me to be broadly representative for the way a large number of linguists within the cognitive/functional camp view the current situation in linguistics. Any serious investigation into the reasons for this schism is beyond the scope of this dissertation, but suffice it to say that different theoretical commitments have led the two theoretical groups to have, at times, very different research interests and research focus and also to rely on different methods for procuring data. Gries goes on to say:

> Labov and Levelt, among others, already showed in the early 1970s that the judgments that were adduced to support theoretical developments were far from uncontroversial and that better ways of gathering judgment data are desirable. Over the last few years, corpus data have especially become one of the most frequently used alternative types of data. This movement towards empirically more robust data is desirable. (Gries, 2014, 16)

In addition to corpus methods, he also mentions psycholinguistic, experimental methods. This focus on non-elicited, naturally occurring data (in the form of corpus data) and experimentally collected data is typical within non-generative frameworks of linguistics, and the use of linguistic intuitive judgements as evidence is often actively discouraged.

Another, probably not too uncommon, attitude towards the use of intuitive judgements as evidence within non-generative linguistics is expressed by Geoffrey Sampson:

> I guess that most English speakers, if asked the grammatical status of the string *John loves Mary*, would say that it was a good English sentence — and I surmise that they would be correct. But, if the point were challenged, the respectable response would be to look for hard evidence, not to assert the clarity of one's intuitive judgement. (Sampson 2001, 125, original emphasis)

I think that these two points of view exemplify some of the most common views on the use of intuitive judgements as evidence in non-generative linguistic frameworks. On the one hand, there is the sense that using intuitive judgements is something that should be avoided in favour of more "empirically robust" methods. On the other hand, there is a pragmatic stance that in many humdrum cases, speakers' intuitive judgements might be good measures of the grammaticality of utterances, and so in easy cases one can get away with relying on them as evidence – but in cases of disagreements, they should be backed up with other kinds of evidence.

Examples of non-generativists using informally collected linguistic intuitive judgements as evidence can, for instance, be found in a recent grammar of Danish by linguistic researchers identifying as belonging to the functional tradition of linguistics (Hansen and Heltoft, 2011). The authors mention their native speaker intuitions as one of the sources of their work.

An example of experimentally collected judgements being used as evidence within non-generative linguistics is found in the exchange between Dąbrowska (2008, 2013) and Ambridge and Goldberg (2008). They use acceptability judgement tasks as one source of evidence to test different usage-based/cognitive hypotheses about sentences with long-distance dependencies. While some long-distance dependencies appear to be acceptable to native speakers of English (e.g., (4) below), others appear to be unacceptable to native speakers (e.g., (5) below).

(4) What do you think you're doing? (Dąbrowska, 2010)

(5) *Who did that she knew _ bother him? (Goldberg 2006, quoted in Dąbrowska 2013)

In generative linguistics, this difference is explained in terms of subjacency (Chomsky, 1973; Newmeyer, 1991). Dąbrowska and Ambridge and Goldberg debate two usage-based/cognitive alternative explanations. Ambridge and Goldberg (2008) argue for an explanation based on information structure that they call "Backgrounded Constituents are Islands". In short, their explanation that (5) above is unacceptable is that the referent ("who") is not part of the potential focus domain of the sentence. Instead, they argue, it is part of a presupposed or "backgrounded" constituent (here, it is presupposed rather than asserted that the referent of "who" bothered "him"). This is not the case in (4).

Dąbrowska (2013) tests and partially replicates the results of Ambridge and Goldberg but also concludes that some results have to be explained by another principle which she calls the Lexical Template Hypothesis. This hypothesis states that speakers apply lexical templates (such as *Wh-word do you think subordinate-clause-with-gap?* for (4) above) when producing, comprehending,

and judging sentences in at least some cases, rather than necessarily applying maximally general rules. In short, these studies show how experimentally collected linguistic intuitive judgements are, at least in some cases and by some authors, used to test non-generative linguistic hypotheses.

4.2.1 Etiology accounts

Since intuitive judgements have a much less prominent position within non-generative linguistics than within generative linguistics, there has been significantly less discussion in the non-generative literature of the justification of the practice than in the generative literature. Dąbrowska (2016) is a notable exception. She discusses the use of linguistic intuitive judgements from a non-generative, cognitive linguistics standpoint (see also Willems 2012 and Gibbs Jr. 2006 on intuitive judgements about concepts and semantics within cognitive linguistics). In this section, I give a brief introduction to her account of the etiology of linguistic intuitive judgements, and I then introduce an account by Luka (2005), another etiology account that is not specifically generative.

4.2.1.1 Speakers' theories about language

Dąbrowska (2016) notes how linguistic methodology has become more rigorous over the last years, and how the practice of using intuitive judgements as evidence has come under scrutiny, e.g., due to worries over experimenter bias. Still, she writes, when studying something as complex as human language, using all available evidence can only be an advantage. Thus, even if more rigorous methods are available for studying at least some aspects of the phenomenon of language, linguistic intuitive judgements may still have a role to play. She advocates using basic scientific methodology to collect and analyse linguistic intuitive judgements, including using participants who are blind to the aim of the study, using large samples of materials and participants, and applying appropriate statistical analyses.

On her account, linguistic intuitive judgements are based on speakers' theories of grammar (and her position thus has some resemblance to Devitt's Modest Explanation as laid out in chapter 2). Dąbrowska (2010) showed that linguists' intuitive judgements differ from the judgements of naive speakers, presumably either due to the linguists having different theories about language or due to the linguists being more frequently exposed (through linguistic literature) to structures that were similar to the ones tested, or a combination of both. On her account, like on Devitt's, the intuitive judgements of the folk are also influenced by their folk linguistic theories of language.

One difference from Devitt's account is that Dąbrowska (2016) focuses on

how both our folk linguistic theories and experts' linguistic theories are shaped by our experience with *written* language. To show how this is the case, she reviews evidence that speakers' intuitive judgements about what is usually taken to be basic linguistic units are heavily influenced by written language norms. The first linguistic unit she looks into is the sentence. Especially within syntax and the practice of using intuitive judgements of morphosyntactic well-formedness as evidence, the sentence seems like a natural linguistic unit. However, Dąbrowska (2016) shows how naturally occurring (non-scripted) speech is not easily and unambiguously divisible into neat sentences but rather contains unintegrated fragments and indeterminate clause boundaries.

She also reports on studies on languages where there is no long tradition of sentence individuation within the written language, as wells studies of English speaking children's use of punctuation marks. These show, in short, that speakers might be entirely competent in their languages without having a clear sense of how to individuate sentences (in the way we are used to thinking of "sentences"). She presents similar evidence about speakers' judgements about what counts as a word.[1] Our idea of what makes a sentence or a word, she concludes, is a product of our particular written tradition in which these are important units rather than an inevitable part of our native linguistic competence.

She also notes how this affects intuitive judgements of grammaticality/acceptability/ specifically (she uses the term "grammaticality judgements"). As one would expect if these judgements were to a large degree shaped by speakers' experience with written language, studies have shown that pre-literate children and illiterate adults are bad a providing grammaticality judgements (she references a study by Karanth et al. 1995). She argues that even literate adults' grammaticality judgements are, effectively, about what is and is not considered acceptable in written language. She supports this by listing constructions which are common in everyday spoken language but which speakers typically judge to be ungrammatical. One example she gives is the following:

(6) There's always one or two do this each year (Dąbrowska, 2016, 68)

This utterance would likely be judged to be ungrammatical by speakers (especially if they are judging by a written language standard), despite the fact that the utterance adheres to a form found quite commonly in spoken language.

She further notes that not just the judgements but even the development of the grammar of speakers is affected by learning to read and write, citing effects independent of education level found among adult speakers. Those speakers who

[1]She also makes a similar point about intuitive judgements about phonemes, which I will not go into as it goes beyond our focus on intuitive judgements of morphosyntactic well-formedness.

read more performed better on grammatical tests than speakers (of comparable education backgrounds) who read less.

She takes all this to show that what we get a measure of when we elicit intuitive judgements is not the speaker's language or linguistic competence "in its 'natural' state" (Dąbrowska, 2016, 70). Rather, the judgements reflect both speakers' linguistic competence and their linguistic theories, both of which are shaped to a large degree by the speakers' experience with written language systems. This does not mean that languages with a long literary tradition should not be studied or that only illiterate speakers should be used as subjects – then we would be ignoring a central aspect of the phenomenon of language. However, it does mean that we should not treat intuitive judgements as direct reports of the speaker's (steady state) native linguistic competence.

In many ways, this account is similar to the one Devitt defends. On both views, linguistic intuitive judgements are based on the theories of language that speakers hold. The two authors draw different conclusions, however. Devitt argues that, to the extent that we should use intuitive judgements as evidence, we should use the judgements of those subjects who posses the best theories, i.e., linguists. Dąbrowska argues that we should use participants who are blind to the purpose of the study, and that since linguists have different theories of language than lay subjects, their intuitive judgements cannot be assumed to be representative of the population (see Dąbrowska 2010).

4.2.1.2 Affective evaluations

Luka (2005) is another example of a treatment of linguistic intuitive judgements from a perspective that is not particular to generative linguistics. She presents what she calls a "cognitively plausible model of linguistic intuition" which, she posits, can be plugged into different theories of grammar. On the model, "implicit memory for previously encountered syntactic patterns contributes to linguistic knowledge" (Luka, 2005, 491), which fits well with usage-based/constructivist theories of linguistics (e.g., Tomasello 2005).

She builds her model on research into implicit learning, specifically of artificial grammars. She references a study by Gordon and Holyoak (1983) in which participant were presented with utterances constructed in accordance with a novel, artificial grammar. Later, those participants displayed higher ratings of liking (affective evaluations) towards new utterances that were produced by that same grammar, generalising across tokens to types, in contrast to comparable strings not consistent with that grammar. She proposes that the affective evaluation is due to perceived ease of processing which is due to the participants' learning of the grammar. This she takes to show that affective evaluation func-

tions as a measure of the speaker's competence with the grammar. Processing will, other things being equal, be effortless in those cases where the sentence is part of the speaker's language, and the speaker will evaluate those sentences positively. In the cases where the sentence is not part of the speaker's language, the processing will fail, and this will make the speaker evaluate the sentence negatively.

According to Luka's proposal, this affective evaluation of a sentence is something which the speaker can consciously access and, e.g., report or make the basis for further deliberations, for instance when asked to make a meta-linguistic judgement about a sentence.

In short, on her model,

> [...] affective evaluations regarding ease of sentence comprehension are one of the mechanisms by which unconscious information about language enters conscious awareness and influences the report of linguistic judgements. (Luka, 2005, 483)

Luka's proposal will be revisited in chapter 8.

Whereas the account presented by Dąbrowska has some resemblance to Devitt's account, the proposal by Luka is more like the competence-based accounts presented by Devitt's critics. On both her and their accounts, the main justification for the use of intuitive judgements as evidence for grammatical theories comes from the claim that the intuitive judgements have a special etiology: They are closely based on something that, fairly reliably if not completely perfectly, tracks the speaker's linguistic competence. On the accounts by Devitt and Dąbrowska, the use of intuitive judgements as evidence in linguistics is justified because (and to the extent that) the speaker's theories of language are fairly accurate.

As far as I am aware, there has been no interaction between the debate of the etiology question in generative linguistics and these non-generative proposals.[2]

4.2.2 Methodology

Like the etiology question, the issue of how to best collect and analyse linguistic intuitive judgements has not received nearly as much attention within the non-generative literature as it has within the generative literature. One non-generative discussion of the use of intuitive judgements as evidence is due to William Labov, who founded the field of variationist sociolinguistics. He proposed the following four working principles for the use of intuitive judgements as evidence in linguistics:

[2] Although Maynes and Gross (2013) mention Luka's proposal in passing.

1. "The consensus principle: if there is no reason to think otherwise, assume that the judgment of any native speaker are characteristic of all speakers of the language."

2. "The experimenter principle: if there is any disagreement on introspective judgments, the judgments of those who are familiar with the theoretical issues may not be counted as evidence."

3. "The clear case principle: disputed judgments should be shown to include at least one consistent pattern in the speech community or be abandoned."

4. "The principle of validity: when the use of a language is shown to be more consistent than introspective judgements, a valid description of the language will agree with that use rather than introspections." (Labov, 1975, 103, 112)

As support for the fourth principle, Labov presents results from his own studies. In spite of producing certain phonetic differences with great regularity in their everyday language use, speakers could not consistently identify the same differences in samples of their own speech in intuitive judgement tasks. In cases like this, he argues, the consistent patterns in the language use should not be ignored because of the lack of consistent intuitive judgements.

Apart from the fourth principle, these principles accord well with the recommendations of the proponents of Experimental Syntax as described in chapter 3. We also saw that Dąbrowska's recommendations for how to collect and analyse intuitive linguistic judgements aligned with the experimentalist stance. I am not aware of any non-generative literature arguing for the traditionalist stance on this question. In section 6.2.2, we will see what the majority of non-generative participants think about these methodological issues.

4.3 Summing up

Although linguistic intuitive judgements have a special role as evidence in generative linguistics, it is also used in other linguistic frameworks, although they seem to be treated as a more defeasible type of evidence outside than within generative linguistics. And while it is not widely debated in the non-generative literature, some proposals for etiology accounts as well as some methodological guidelines can be found within non-generative linguistic literature. A justification for the evidential use of linguistic intuitive judgements should take this into account, either by explicitly narrowing its focus to one framework or by providing a justification that works across different linguistic frameworks.

Part II

Questionnaire study

Chapter 5

Study design, materials, and participants[1]

In this chapter, I present the design of a questionnaire study meant to investigate the current views among linguists on the two debates presented in chapters 2 and 3: the etiology debate and the Experimental Syntax debate. I first lay out the purpose of the study and its overall structure, and then I go into some details of the questionnaire design. I then present the main questions of the questionnaire, and afterwards I present some background information about the participants in the study. Finally, I consider some potential weaknesses of the design of the study. The results of the study are presented in chapter 6.

5.1 Purpose

The purpose of this study is to investigate how linguists currently view the questions of etiology/justification and Experimental Syntax, which have been brought up in the debates outlined in chapters 2 and 3.

More specifically, the main questions that the study seeks to answer are the following:

- Do a majority of generative linguists hold the Voice of Competence view (VoC), as characterised by Michael Devitt, regarding the use of intuitive judgements as evidence in linguistics? Or, do a majority hold any of the other views presented in the debate (the Modest Explanation and the Critics' Views), or some other view entirely?

[1] In accordance with the Graduate School of Science and Technology rules, parts of this chapter was used in the progress report for the qualifying examination (specifically, section 5.1, 5.2, 5.3, and 5.6 of this chapter). These sections have been slightly updated.

- Do a majority of non-generative linguists hold the Voice of Competence view or any of the other views presented in the debate?
- What are the majority positions within generative and non-generative linguistics respectively about the issues of the Experimental Syntax debate?

Some other questions that the study hopefully will be able to shed light on as well are:

- Is there a difference between generativists and other linguists when it comes to their views regarding the etiology of linguistic intuitive judgements and the Experimental Syntax debate?
- Is there a difference between experimentalists and traditionalists (see chapter 3) when it comes to their views regarding the etiology debate?
- Is there a difference in how linguists with different specialisations within linguistics view the issues of the etiology debate and the Experimental Syntax debate?

5.2 Overall structure of study

In order to answer the questions outlined above, a mixed method study was carried out. The data comes from a questionnaire, which was constructed based on the debates on etiology/justification and Experimental Syntax in the literature (see chapters 2 and 3). The main results from the study come from a quantitative analysis of the answers from the questionnaire. Some of the patterns found in this analysis are explored in more detail through a qualitative analysis of open-ended comments provided in the same questionnaire.

Morgan (2014, 155) distinguishes three ways a qualitative study can extend on the findings of a quantitative study:

1. explore the data to "produce broad insights into how and why a particular set of results occurred",
2. investigate and "explain specific patterns in the quantitative data", or
3. illustrate key results from the quantitative study.

In this study, the qualitative analysis of the comments is meant to follow the second option: to shed more light on specific patterns found in the quantitative analysis. For instance, in cases where there is no significant majority for any of the given options, doing a qualitative analysis of the comments allows us to dig into the motivations for and reservations about the different positions, i.e., to explore potential explanations for the lack of a "clear" result.

The quantitative analyses are done with standard frequentist statistic tests (using nominal categories), and the qualitative analyses are done in the form of qualitative content analysis. The methods and particular tests are introduced and explained as they become relevant throughout chapter 6.

5.3 Questionnaire design

In this section, I give an overview of the main considerations going into the design of the questionnaire.

5.3.1 Defining and operationalising concepts

As one of the aims of the questionnaire is to investigate the majority view on the justification question among linguists, a number of the questions that make up the questionnaire are based on the areas of disagreement in the etiology debate (as described in chapter 2). The questions thus reflect that the main discussion of the etiology of linguistic intuitive judgements has taken place within the generative linguistics framework.

Each of the three views (VoC, the Modest Explanation, and the Critics' Views) is taken to be one concept. Each issue that the parties disagree about is taken to be a variable within each concept. The different stances that are found in the literature regarding each issue/variable are taken to be the categories of that variable. For example, one of the questions that divides the sides of the debate is whether linguists should use intuitive judgements of acceptability or intuitive judgements of grammaticality as evidence for their theories. This issue (acceptability or grammaticality) is a variable within each concept (VoC, the Modest Explanation, and the Critics' Views), and the categories of that variable are "acceptability" and "grammaticality".

Table 5.1 summarises what categories are selected on each variable for each of the three concepts (using shorthand names for each variable). The question numbers in parentheses (Q...) show what questions in the questionnaire each of the variables correspond to. See the phrasing of the questions in section 5.4 below (or in appendix A).

Variables / Concepts	Modest Explanation	VoC	Critics' Views
Origin (Q2)	Reflections	Competence	Competence
Acceptability/grammaticality (Q3)	Grammaticality	Acceptability	Acceptability
Deduced (Q4)	Not deduced	Deduced	Not deduced
Fallibility (Q5)	Fallible	Infallible	?
Subject matter (Q6)	Linguistic patterns	Mental capacities	Mental capacities
Rules (Q7)	Not implemented	Implemented	?

Table 5.1: The three positions

As can be seen in the Critics' Views column of the table, and as described in chapter 2, the critics sometimes disagree on certain variables, and this should be reflected in the analysis of responses.

The variables and their categories are the bases for the formative indexes described in section 5.3.2 below.

In a similar way, the Experimental Syntax debate was broken down into a number of issues. These were the following.

- Does the traditional method provide good evidence for grammatical theories or should we use experimental methods?
- Are linguists better subjects than naive, ordinary speakers?
- Is reliable data more important than theoretical virtues?
- Does the gradience found in experimental studies most likely come from a graded grammar or from extra-grammatical factors?

In section 5.4 below, I show how these issues were phrased in the questionnaire.

5.3.2 Formative indexes

An index is a standard tool in questionnaire research. It helps researchers investigate complex constructs, and at the same time it reduces some of the problems of validity and reliability that can affect quantitative studies. The general idea behind this tool is to take a complex construct that one is interested in, say, whether a person has a healthy lifestyle, and break that down into several parts, e.g., whether or not they get enough sleep, whether or not they exercise etc. Now, instead of asking just one question ("Do you have a healthy lifestyle?"), you ask a question for each of the aspects or variables that you have broken the complex construct into. From the answers to each of the questions, one then computes the answer to the overall question one wants answered (more details on that below).

In this study, the views on etiology described in chapter 2 and their component parts form the basis of three indexes. Instead of asking participants one question, such as "Do you subscribe to the Voice of Competence view regarding the evidential use of intuitive judgements in linguistics?", they are asked several questions, each of which covers just one aspect of the three views that are under investigation.

The individual questions in the questionnaire are called the items of the index. The indexes in this study are built with so called formative items (Diamantopoulos and Winklhofer, 2001). This means that the components that go

into the index are seen as individual aspects of the construct in question, and they are seen as collectively constituting that construct (e.g., the concept of a healthy lifestyle, or one of the three views on the etiology question investigated here).[2] When one collects data with a formative index, the results for each item are not necessarily expected to intercorrelate as some criteria for the construct may be fulfilled while others are not (this is in contrast to reflexive scales).

According to Petersen (2010), the kind of indexes described here are used in survey studies as they often give more valid measurements than using single questions, i.e., they may make it more likely that you are measuring the thing you actually want to measure (if well-constructed). The idea is that if participants are given a number of questions about specific aspects of a complex construct instead of one question about the complex construct in general, there is a greater chance that they will actually understand and answer the questions in the way they were intended to be understood.

Compared to asking just one question about a complex construct, using indexes furthermore potentially provides benefits for the reliability of a study. This is because asking a number of questions instead of only one means that the random variation in participants' answers is distributed over the multiple questions. If the variation is really random, and not systematic due to some unconsidered variable, this means that the overall negative effects are diminished, and the result is more reliable.

When creating a formative index, all aspects of the modelled concept need to be covered, and each aspect should be included only once. The selection of aspects to include as items relies primarily on the theory/literature, and one should be very careful with the selection process as it cannot be tested statistically whether an aspect/item belongs to the concept or not (this can be done for reflexive scales, on the other hand).

The indexes of this study are summative, which means that for each item in the questionnaire, the scores are added up to form the total score. In the example used above, this would mean that instead of answering the question "Do you have a healthy lifestyle?", participants would indicate their position on several items that together form the complete construct, and those answers would then be summed to form the total score for the construct of "having a healthy lifestyle". There are, however, other ways to calculate the scores for indexes as well, see Hellevik (2002). The total score of an index can primarily be interpreted in relation to something else, e.g., the score of one person as related to the score of another person or as related to the total possible score.

[2] Another type of index can be made with so called reflexive items, and those "indexes" are usually called scales (see Spector 1992). For reflexive scales, the items are seen as tracking something that is the result of the complex construct, not tracking constituent parts of the construct. For instance, being tired often could be seen as the result of an unhealthy lifestyle.

The three indexes of this study were constructed to map the three views mentioned in chapter 2. That is, each view is turned into a separate index. Participants will thus get a score on each of the indexes, and the one they score the highest on will be interpreted as the view they agree with the most. The questionnaire contains six items which are all part of the three indexes. A number of additional items are included to cast light on the Experimental Syntax debate, and there are questions pertaining to participants' backgrounds as well (see the full questionnaire in appendix A). See an overview of how the items build into the indexes in table 5.1 above.

5.3.2.1 Overlapping indexes

The three indexes used in this study have a certain degree of overlap between them (see table 5.1 above). By this, I mean that for some (most, in fact) of the questions in the questionnaire, choosing one particular response will result in getting a point on more than one of the indexes. The literature that specifically deals with overlapping formative indexes is sparse. Based on the literature on overlapping reflexive scales, I have identified some potential problems. Here, I will only mention the two that seem the most problematic (see Spector 1992 and Nicholls et al. 1982 for more discussion of overlapping reflexive scales).

The first problem that I want to mention is that when different sub-scales correlate with another criterion variable, it will not be possible to determine whether the similar relation between the two scales and the variable are because both scales are in fact related to the third variable or whether it is because the scales overlap. As an example, say that two of the views look like they are related positively to, for instance, the participants' geographical location. In that case, we will not know whether that is because they are each actually related to location or because the two indexes share some items. However, the constructs in this case, the three views on the etiology question, overlap themselves. In chapter 2 we saw that on several of the areas of disagreement (variables), two of the views would have the same component view. That is, the three overall views are not different because *all* their component parts are different, instead they are different because they each have different configurations of component views that are shared with the other views. I think the indexes need to reflect this fact although it makes the interpretation of the results less straightforward.

The second problem is that with overlapping scales, the score for one construct may be inflated because of answers resulting from other constructs. Using an example from Spector (1992, 40), I want to argue that this is mainly a problem for reflexive scales, not necessarily for formative indexes. Suppose we have two constructs, anger and anxiety, and that we have built a reflexive scale for

each of the constructs with a number of items on each, corresponding to the aspects of the construct that we want to investigate. "I feel tense" is a component of both scales. Assume that the participant who fills out the questionnaire is angry, but not anxious. Their positive response to "I feel tense" will falsely inflate their score on the anxiety scale.

This would be problematic for a reflexive scale, but for a formative index the components are independent of each other, so the fact that the participant responds positively to "I feel tense" should add to the scores for both constructs.

In conclusion, I will proceed with the indexes presented above even though there are overlaps between them.

5.3.3 Phrasing items

Items have been phrased carefully and undergone a number of revisions in collaboration with colleagues. I have attempted to follow the guidelines from Hansen (2010) and Olsen (2006). These guidelines include the following advice: Minimise the range of possible interpretations for the participant. Use words that are simple, frequent, short, and unambiguous. Avoid implicit assumptions, and avoid asking multiple things at once. Avoid using superlatives and comparatives if it is not made explicit what the thing being compared to is. Also avoid quantifying adverbs like "often" which may invoke different expectations from different participants, use specific terms like "daily" or "weekly" instead. As we shall see in chapter 6, there are still some issues of phrasing left in the final version of the questionnaire.

The term "syntactic intuitions" was used throughout the questionnaire to refer to the kind of linguistic intuitive judgements that are under investigation. This term was chosen both to focus participants' attention on judgements of morphosyntactic well-formedness and to avoid the terms "acceptability" and "grammaticality" as one question in the questionnaire deals explicitly with this distinction.

5.3.4 Ordering items

The order that items are presented in can frame how subsequent items are interpreted (Olsen, 2006). One way to deal with this is to randomise the order of the questions between participants. Some considerations speak against randomising questions, however. One such consideration is the level of attention of the participants. It can be beneficial to have opening questions that catch the interest of the participants and that are easy to answer, and it might be best not to place the questions that are deemed to be the hardest to answer last in the questionnaire as participants might get tired during the process.

The order of questions was not randomised in the present questionnaire. This was partly due to the considerations mentioned above, and partly because I felt that some questions might appear more basic to participants than other questions, and so if the more basic questions were not asked early on, participants might have felt that those questions were missing, potentially influencing their answers to other questions. For instance, if a participant feels strongly that only *acceptability* judgements can be used as evidence, they might find it difficult to answer questions about the use of intuitive judgements as evidence that do not specify whether we are talking about judgements of acceptability or grammaticality. This kind of dependency between questions is hard to avoid, and so the questions that I thought would seem most fundamental to participants were placed early on, so that hopefully participants would answer subsequent questions without worrying about those potentially more fundamental questions.

5.3.5 Item type, format, and answer categories

With a few exceptions, the questions in this questionnaire are what Olsen (2006) calls *opinion* questions, compared to factual questions. More specifically they are opinion questions of the *cognitive* type as opposed to affective or conative opinion questions (i.e., "what do you think?" rather than "what do you feel?" or "what would you do?"). The bulk of the questions in this questionnaire are about the participants' opinions *in general*, not about what they would think about a specific case.

The variables of the discussed views in general have two answer categories for the reason that, in general, two or more different potential positions are discussed in the literature for each component part of the main three views. Therefore, a multiple choice format was chosen for most of the items to find out which of the options participants agree with (if any). For each multiple choice item, there are three or four answer categories: the two or three answers that are of interest in the debate and a third option that participants were asked to fill in themselves if neither of the presented options fit them. A couple of questions were answered by a 5-point Likert scale where participants are asked to indicate their level of agreement with a statement.

The answer categories for the multiple choice items are nominal, and the Likert-scale responses were treated as nominal (Agree, Neither agree nor disagree, and Disagree) in the analysis despite in principle being ordinal. The choice between multiple choice and Likert-scale format for the individual items was mainly driven by whether the issue could be stated as a statement with which participants could agree or disagree or whether several separate answer

options had to be specified.

The open fill-in option is needed in the multiple choice items as we are interested in finding out what view participants hold, not just which of the presented options they find more appealing. For this reason, the prompt for each item is phrased "Please choose the option that best expresses your opinion", and participants were encouraged to use the fill-in option if that best fits their view. Including an "other" option is also mentioned by Olsen (2006) as an important tool to help avoid that participants "make up" opinions on the spot just to comply with the request for an answer.

5.3.6 Instructions

The questionnaire was prefaced with a call for participants containing the link for the questionnaire and, once participants followed the link, a set of instructions. In the call for participants, it was specified that linguistic researchers of any theoretical background who think that linguistic intuitive judgements may be used as evidence for theories of grammar are welcome to participate. Participants were informed that their answers would be treated anonymously. See the call for participants and the instructions in appendix A.

5.3.7 Pilot test

Before sending out the final version of the questionnaire, a pilot test was run on a previous version. The purpose of the test was to find out:

- whether the individual items were understood in the intended way,
- how long it took participants to fill out the questionnaire, and
- whether the questionnaire was missing questions on any particular aspects of the evidential use of intuitive judgements in linguistics.

Five participants (all linguists, and all either university faculty or PhD students) filled out the questionnaire and were subsequently interviewed about their interpretations of the questions.

The main aim of the test was to see whether the questions would be interpreted by participants in the way intended. Here is one example: One question relates to the role of the speaker's mental grammar in the etiology of linguistic intuitive judgements. Participants were asked to complete the sentence "Syntactic intuitions are ..." by choosing one of the two options "... tacitly deduced from the speaker's mental grammar" or "... *not* tacitly deduced from the speaker's mental grammar" (emphasis included) or alternatively to supply another answer themselves. In the interview with one participant, it became apparent that they

understood the question to be about whether linguistic intuitive judgements are affected by noise (from, e.g., performance factors) or not. This is not what this question was intended to cover, so this response was very informative. In the final version, this question was rephrased and now occurs in the form of question 4 (see section 5.4 below).

Other questions were rephrased as well to make the intended distinctions stand out more clearly to participants. One question was dropped as it was highly unclear to participants and not as central to the debate as previously assumed.

5.4 Materials

Below, the questions from the main part of the questionnaire (excluding the background questions) are presented in the order they appeared in the questionnaire (including any typos or other errors that might also have been included in the questionnaire as presented to participants). The boldface text are shorthand names for each question that will be used throughout the next few chapters. They were not displayed in the questionnaire. The numbers refer to the order of the questions in chapter 6. The questions will be presented again in chapter 6 as the respective answers are analysed. See all of the questions, including background questions, in the questionnaire in appendix A.

Stand alone (later Q1)
When studying grammatical phenomena, syntactic intuitions ...

1. ... can in some cases stand alone as evidence.
2. ... can in some cases serve as evidence but can never stand alone.
3. ... can in no way be used as evidence.
4. [Supply other answer]

Subject matter (later Q6)
When I study grammatical phenomena, I ultimately seek to understand ...

1. ... the systematic patterns found in linguistic behaviour.
2. ... the linguistic capacity of the mind.
3. [Supply other answer]

Acceptability/grammaticality (later Q3)
Most native speakers of English would find the sentence 'the dog that the woman that the man saw owned ran' to be an intuitively

unnatural sentence in English. In the terminology of generative linguistics, the sentence is *unacceptable* to those speakers (this is not to be confused with whether or not the sentence is infelicitous to the speakers in some specific context). Most native speakers would also most likely find the sentence intuitively *ungrammatical*. However, some experts argue that it is, in fact, grammatical.

To the extent syntactic intuitions can serve as evidence for theories of grammar, only those syntactic intuitions can serve as evidence that are ...

1. ... acceptability intuitions.
2. ... grammaticality intuitions.
3. [Supply other answer]

Origin (later Q2)
When syntactic intuitions are reliable as evidence, this is *mainly* because ...

1. ... they are speakers' *reflections* about language use, and speakers are to some degree reliable judges about this.
2. ... they express speaker's *competence* in their native language.
3. [Supply other answer]

Rules (later Q7)
Syntactic intuitions is one type of evidence used by linguists to characterise how languages are structured. Let's call the rules that describe this the *structure rules* of particular languages.

The structure rules that linguists describe are sometimes said to be "implemented in the minds of speakers".

1. It is a good hypothesis that structure rules are actually implemented in the minds of speakers.
2. From the structure rules we observe, we can only infer that the mind works *as if* it was following those rules.
3. From the structure rules we observe, we cannot infer anything about how the mind processes language.
4. [Supply other answer]

Implementation (later Q8)
There must be something in the mind that gives rise to what we call "rules of grammar".

1. The rules of grammar are probably explicitly represented in the mind. If one could look into subjects' minds, one could find explicit rules.
2. The rules of grammar are implemented in the mind, but they are probably not explicitly represented.
3. The rules of grammar are probably not implemented in the mind.
4. [Supply other answer]

Deduced (later Q4)
Syntactic intuitions are sometimes said to be "deduced from the speaker's mental grammar".

1. This is probably a poor description of how intuitions are formed.
2. This is a good way to talk about how intuitions are formed but should probably not be taken too literally.
3. This is likely to be the actual process of how syntactic intuitions are formed in the mind.
4. [Supply other answer]

Fallibility (later Q5)
Imagine you could abstract away all performance factors that might influence a speaker's syntactic intuitions. In that case, it would be possible for the resulting syntactic intuitions to be mistaken about the grammatical properties of the sentence.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

Experts and ordinary speakers (later Q9)
The syntactic intuitions of linguists working on theories of grammar are, *on average*, ...

1. ... worse evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
2. ... equally good evidence for theories of grammar as the syntactic intuitions of ordinary speakers.
3. ... better evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
4. [Supply other answer]

Traditional methods (later Q12)
In general, consulting one's own or one's colleague's syntactic intuitions produces good evidence for theories of grammar.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

Experimental methods (later Q13)
In general, syntactic intuitions should be collected and analysed by experimental methods from large numbers of speakers and using statistical tests.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

Theoretical virtues (later Q10)
In general, if one has to choose, it is more important that a theory is built on reliable data than that it lives up to theoretical virtues such as simplicity, elegance, and fruitfulness.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

Gradience (later Q11)
Some linguists use gradient rather than binary scales to collect syntactic intuitions.

1. Well-designed gradient scales may very well reflect real degrees of grammaticality.
2. Even well-designed gradient scales probably just capture effects that are not due to grammaticality.
3. [Supply other answer]

Significance 1 (later Q14)
Can you think of a situation, before answering this survey, where you thought about whether syntactic intuitions can serve as evidence for theories of grammar?

1. Yes
2. No
3. [Supply other answer]

Significance 2 (later Q15)
Linguists who use intuitions as evidence should set aside time to consider *why* intuitions can serve as evidence for their theories.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

After each question, participants had the optional possibility of commenting on the topic of the question. This prompt had the following form:

Optional:
Would you like to elaborate on your answer or comment on anything else regarding this section?

After the main section of the questionnaire, participants were asked a number of questions about their demographic background and previous use of intuitive judgements as evidence for grammatical theories. At the end of the questionnaire, participants could sign up for updates about the study, indicate that they would be interested in participating in potential follow-up interviews (which, in the end, were not performed), and they were also able to sign up for a lottery for 10 vouchers for Amazon.com for $25 each.

5.5 Participants

The questionnaire was distributed through LinguistList in early July 2017, and data collection continued until mid September 2017 when it was decided that a sufficient number of responses was reached, and the stream of responses had slowed down considerably. There were 239 responses in total to the questionnaire. Of those 239 responses, 47 completely empty responses were excluded, and another 47 were excluded because the participants did not finish the questionnaire. Finally, of those who finished the questionnaire, 11 were excluded because they did not live up to the inclusion criteria which were a) being enrolled in or having a PhD ($n = 6$ excluded), b) believing that linguistic intuitive judgements can in any way serve as evidence ($n = 4$ excluded), and c) mainly doing research within linguistics or closely related fields ($n = 1$ excluded). This leaves us with $n = 134$ responses.

Information about the participants' academic position, age, geographical region, English proficiency, and more was collected through a background section presented at the end of the questionnaire. Participants also answered a question about their theoretical orientation, choosing from the following options: a) formal/generative, b) functional/cognitive, c) other/atheoretical, or d) mixed. If they answered (c) or (d), they were asked to elaborate. This information was used to divide participants into two groups. In the one group we have participants who described their theoretical perspective on language as formal/generative (a; $n = 57$) or mixed (d) and mentioned formal and/or generative linguistics as part of their theoretical influences ($n = 16$). In the other group we have all other participants: $n = 37$ who chose functional/cognitive (b), $n = 12$ who chose other (c), and $n = 12$ who chose mixed (d) but did not mention formal/generative linguistics as one of their influences. Among those who chose other, the most commonly mentioned theoretical orientations were functional linguistics, and, separately, cognitive linguistics, usage-based linguistics, and descriptive/atheoretical linguistics. Among those non-generative participants who chose mixed, many mentioned structural linguistics as an influence, while others mentioned sociolinguistics/variationism, corpus linguistics, and statistical/computational linguistics.

Table 5.2 (page 86) presents an overview of the background information for the generative group and for the non-generative group respectively.

Participants were also asked about their own use of linguistic intuitive judgements as evidence for grammatical theories. Out of the generative group ($n = 73$), 70 (96%) reported having used linguistic intuitive judgements in their own research while 3 (4%) reported not having done so. In the non-generative group ($n = 61$), the corresponding numbers were 42 (69%) and 19 (31%) respectively. The participants who reported having used linguistic intuitive judgements in their own research were further asked about whether they had used their own linguistic intuitive judgements or those of other informants, and whether they had used linguistic intuitive judgements collected informally (in the armchair), through fieldwork, or through surveys. The results are presented in table 5.3 (page 86; note that the numbers do not sum to the total number of participants in each group as each participant was allowed to select multiple answers). Participants also had the option to specify if they had used linguistic intuitive judgements collected in any other way. Participants who chose this option mentioned using linguistic intuitive judgements that they had gotten from others' publications and linguistic intuitive judgements collected in behavioural experiments.

	Non-gen ($n = 61$)	Gen ($n = 73$)
30 or under	13 (21%)	12 (16%)
31-40	19 (31%)	28 (38%)
41-50	11 (18%)	17 (23%)
51-60	5 (8%)	10 (14%)
60 or over	13 (21%)	6 (8%)
Total	*61 (100%)*	*73 (100%)*
Male	30 (49%)	39 (53%)
Female + Other	31 (51%)	34 (47%)
Total	*61 (100%)*	*73 (100%)*
PhD student	16 (26%)	15 (21%)
Researcher	45 (74%)	58 (79%)
Total	*61 (100%)*	*73 (100%)*
Intermediate	4 (7%)	0 (0%)
Advanced English	32 (52%)	36 (49%)
Native English	25 (41%)	37 (51%)
Total	*61 (100%)*	*73 (100%)*
US or Canada	9 (15%)	26 (36%)
Europe	40 (66%)	36 (49%)
Rest of the world	12 (20%)	11 (15%)
Total	*61 (100%)*	*73 (100%)*

Table 5.2: Participants' backgrounds

	Own	Others'	Armchair	Fieldwork	Surveys	Other
Gen ($n = 73$)	59 (80%)	69 (95%)	63 (86%)	39 (53%)	52 (71%)	6 (8%)
Non-gen ($n = 61$)	24 (39%)	36 (59%)	21 (34%)	14 (23%)	22 (36%)	5 (8%)

Table 5.3: Participants' use of linguistic intuitive judgements (note that participants could select more than one option)

Finally, participants were asked about their area of specialisations within linguistics and were asked to choose 1 to 3 options. See the options and the number of participants who chose each option in table 5.4 (page 87). Note that the total number of specialisations chosen sums to more than the total number of participants as each participant could choose more than one specialisation. 14 participants chose to specify additional specialisations, see an overview of

those in table 5.5.

Specialisations	Non-gen ($n = 61$)	Gen ($n = 73$)	Total
Morphology	8	23	31
Syntax	16	62	78
Phonetics	9	2	11
Phonology	9	9	18
Semantics	14	29	43
Pragmatics	13	12	25
Sociolinguistics	16	1	17
Anthropological ling.	5	1	6
Psycholinguistics	17	16	33
Cognitive linguistics	25	1	26
Typology	8	4	12
Historical linguistics	10	8	18
Theoretical linguistics	5	29	34
Language acquisition	13	18	31
Text linguistics	3	3	6
Corpus linguistics	26	9	35
Other	9	5	14

Table 5.4: Participants' specialisations, by theoretical group (note that participants could select up to three specialisations)

Other specialisations	Total
Computational(/mathematical) linguistics	4
Linguistic impairment	1
Grammar	1
Literary Linguistics	1
Discourse analysis	1
Syntax-semantics interface	1
Contact linguistics	1
Applied linguistics	1
Language teaching	1
Philosophy of language	1
Linguistic variation	1

Table 5.5: Other specialisations

5.6 Potential weaknesses of the design

One potential weakness of the design of this study is that each variable is only represented in the questionnaire through one item. Alternatively, several items for each variable could have been used to calculate a mean score for each variable (this is done to heighten reliability as random variation is then spread out

across a number of items). This mean score could then have been used in the summative indexes. The reasons for not doing this were to keep the questionnaire short and because of difficulties with phrasing these complex conceptual statements in unambiguous and understandable ways for even just one item per variable.

A second, and related, weakness of the design is that items are in general phrased in one direction. Take, for instance, a question about your opinion about a brand of coffee. Items can be phrased positively, e.g., "I think it tastes good", or negatively, e.g., "I think it is too expensive". Both the negative and positive versions can be seen as contributing to answering the question of how much the participant likes this brand. To add up the answers so that a positive reaction to the first example and a negative reaction to the second example would be counted as two positive answers, one would have to inverse the scales used for, say, the negative items to fit the answers given on the positive items. Again, because the items were already difficult to phrase clearly and unambiguously in accordance with the guidelines mentioned above, this approach was not taken.

There is another potential weakness in the way the formative indexes are constructed. As there are six variables and so six questions making up the indexes, even in case we get clear majorities for each question, it might potentially be the case that the result is a draw between the three views (things get a bit more complicated, though, because the choice of categories will load more than one index for most variables). This is maybe less problematic than it might seem, however, as in that case we would have a majority view that is clearly different from each of the three views in the debate, and such a result should be illustrated by the way the indexes are loaded. Also, the construction of the indexes mainly rests on the issues debated in the literature, so one could not easily have added or removed another variable. In sum, this potential weakness is a necessary part of the design.

Some of the choices made about what answers to allow participants to choose made the analysis more difficult than it might have needed to be. One example is that in the background section, participants were in some cases allowed to check more than one option that applied to them (e.g., in the case of specialisation within linguistics, participants could choose multiple options). This choice was made so as to allow participants to give as much information as possible and to portray themselves as accurately as possible (it might be hard to choose just one area of specialisation). This design choice made the analysis of potential associations between answers to other questions and specialisation in linguistic much more complex than it would have been if participants had only been asked to choose one main area of specialisation as tests had to be run for each specialisation (see section 6.3). The higher number of tests also raises the risk

of type 1 errors, i.e., the risk of rejecting a true null-hypothesis.

Another case in which the analysis would have been easier if different design choices had been made is the case of the open "Supply another answer" options to each of the multiple-choice questions. While these comments do provide valuable insight into participants' thinking on the issues in question, leaving out this option would have made the analysis of each question simpler, and in the post-hoc tests (see section 6.1.1) fewer comparisons would have had to be made, which would have made a difference for the Bonferroni-corrections applied. This, in the end, would potentially have lowered the risk of type 2 errors, i.e., the risk of failing to reject a false null-hypothesis. In both these cases, the consideration of participants' possibility to express their views as accurately as possible was given more weight than considerations of the statistical analysis. The "other" option also allowed participants who did not have an opinion on particular issues to "opt out" of answering, instead of semi-randomly choosing one of the provided answers.

In general, it can be a problem to rely on participants' self-reports. For instance, if asking about a particular behaviour, one is dependent on participants being able to recall and report that behaviour accurately. In this questionnaire, since we are interested in participants' opinions rather than behaviour, accurate recalling hopefully should not be an issue. Participants might, however, also for some reason be hesitant to accurately report particular behaviour or opinions. This risk if hopefully minimised by the fact that the issues that are covered in this questionnaire are not particularly sensitive and by the fact that participants were informed at the opening of the questionnaire that their responses would be treated anonymously.

Finally, the participants self-selected for the survey rather than being randomly selected from the relevant population of linguists. This means that the sample is not representative (it is what is sometimes called a convenience sample). The target population is linguistic researchers who believe that linguistic intuitive judgements can serve as evidence in linguistics. Since participants self-selected for participation by responding to a call on LinguistList, we cannot say anything about to what extent this sample is representative of the target population. LinguistList was used for lack of a good resource listing linguistic researchers from which a random selection could have been made.

Overall, these weaknesses mean that the results we get in this study need to be seen as a first approximation of an answer to the questions that motivated the study in the first place. The results, hopefully, give a sense of what the answers might look like, but for more firm conclusions one would need to see the results replicated, preferably in studies that avoid some of the design weaknesses that are part of this study.

Chapter 6

Results

In this chapter, I present the results of two types of analyses of the data obtained through the questionnaire presented in chapter 5. In sections 6.1 and 6.2, I lay out what the majority answer to each question in the questionnaire was (using chi-square Goodness of Fit tests). I first present the results from the generative group of participants and then the results from the non-generative group of participants. In section 6.3, I investigate potential associations between answers to individual questions in the questionnaire and other variables such as specialisation and theoretical orientation (using Fisher's exact test).

An alpha-level of .05 was used for all statistical test, excluding the post-hoc tests where Bonferroni corrections were applied (Shaffer, 1995).

First, I will briefly summarise the results. Recall that the main aims of the study was to investigate what the majority view is on the etiology debate and the Experimental Syntax debate within generative and non-generative linguistics respectively. Within the group of generative participants, there was not a majority of participants behind any of the three views of the etiology debate outlined in chapter 2 (i.e., the Voice of Competence view, the Modest Explanation, and the Critics' Views). The view that did emerge as the majority view on the etiology question among generative linguists in this survey can instead by characterised as follows: Linguistic intuitive judgements can serve as evidence for grammatical descriptions because of their connection to the speaker's linguistic competence, and if we could filter out the influence of performance factors from intuitive judgements, they would be infallible about the correct grammatical status of sentences (though I will discuss some reservations about this result later). However, the majority of generative participants did not believe that intuitive judgements are "deduced" from speakers' mental grammar or that the structure rules of speakers' languages are necessarily implemented

in speakers' minds.[1]

Overall, the view that emerges as the majority view on the etiology question within the non-generative group is quite close to the Modest Explanation with the exception that it is speakers' judgements of acceptability, rather than their judgements of grammaticality, that should figure as evidence in linguistics. On this view, linguistic intuitive judgements can be used as evidence for grammatical descriptions because they express speakers' reflections about linguistic behaviour. The judgements that should be used as evidence are judgements of acceptability rather than grammaticality, and the ultimate goal of grammatical research is to study the externally observable patterns in linguistic behaviour. On the majority non-generative view it is not the case that structure rules are implemented in the minds of speakers, intuitive judgements are not deduced from the rules of speakers' mental grammars, and linguistic intuitive judgements would not give us direct, infallible access to truths about speakers' languages if only we were able to filter out performance factors.

As for the Experimental Syntax debate (see chapter 3), we did not see clear-cut support for one side over the other among the generative participants. The non-generative participants, on the other hand, agreed more with the experimentalist than with the traditionalist side of the debate.

6.1 The generative majority view

In this section, I present the results from the large generative group (n = 73).[2] The general idea is to go through each question and find out whether one answer to that question was given significantly more frequently than the rest of the options and then to put together the majority view based on those results.

First, I go through the analysis of one example in considerable detail to illustrate how the rest of the analyses were performed. If the reader is familiar with the method of analysis, this section can be skipped or skimmed, or it can be consulted later as a reference for how a particular result was obtained.

6.1.1 Example analysis

At the opening of the questionnaire, participants were asked to complete the following sentence by choosing the option that expressed their opinion the best:

[1] The paper Brøcker (forthcoming) is based on these results (presented in section 6.1.2) and the parts of chapter 2 that outline the Voice of Competence view.

[2] This group consists of both those participants who exclusively self-identified as formal/generative (referred to in the following as "the small generative group", n = 57) and those that mentioned formal/generative linguistics as one of more influences (n = 16), see section 5.5. I will note if and how the results of the small generative group differed from the reported results from the large generative group.

Question 1: Stand alone

When studying grammatical phenomena, syntactic intuitions ...

1. ... can in some cases stand alone as evidence.
2. ... can in some cases serve as evidence but can never stand alone.
3. ... can in no way be used as evidence.
4. [Supply other answer]

The idea behind this question was a) to be able to exclude participants who did not think intuitive judgements can in any way serve as evidence in linguistics (as this was one of the inclusion criteria for the study, see section 5.5), and b) to find out whether participants view intuitive judgements as a type of evidence that can be used on its own, or whether they think it can mainly function as supporting evidence alongside other evidence.[3] Recall that in the questionnaire, the term "syntactic intuitions" is used to refer to the linguistic intuitive judgements of morphosyntactic well-formedness that are the focus of this dissertation (see section 5.3.3).

The analysis showed that option 1, *can in some cases stand alone as evidence*, was significantly more frequent than the other options within the large generative group. I will go through the analysis of the answers to this question in detail as an illustration of the analysis done for all the questions in sections 6.1 and 6.2.

As those participants who chose *can in no way be used as evidence* were excluded from the study, just three options remain. See the distribution of answers (excluding option 3) in table 6.1. 1 participant chose *Supply other answer*. This participant wanted to express a stronger commitment than those expressed by the given options and supplied the answer "Can always stand alone as evidence".[4]

Answer	In some cases	Never alone	Other
Frequency	64 (88%)	8 (11%)	1 (1%)

Table 6.1: Stand alone (Q1), distribution of answers, generative group

A chi-square Goodness of Fit (GOF) test was performed to test the null-hypothesis that the three options are equally distributed (i.e., chosen equally often, disregarding random noise). In practice, this would mean that with 73

[3]The pilot study suggested that a binary set of options was too simple, and that it was important to give participants the option to make the reservation that option 2 expresses.

[4]Recoding this answer to *can in some cases stand alone as evidence*, quite intuitively, does not change the outcome of the analysis as it merely adds more distance between the most frequent answer and its closest competitor.

participants, the three options would be chosen by roughly 24 participants each. The test showed that the options were not equally distributed, $\chi^2(2, n = 73) = 98, p < .001$, and the null-hypothesis is rejected. The alternative hypothesis, which is that there is a difference in the distribution of the answers, is accepted. The effect size as measured by Cramer's V is large with $V = .82$.[5] Informally put, statistical significance is a measure of how likely a result is to occur by chance, and effect size is a measure of how big or small a result is. Statistical significance and effect size are independent, as an observed difference can be statistically significant but small, and an observed difference might be large but statistically not significant.

All we know at this stage is that there is a significant difference in frequency between some of the answers and that this difference is comparatively large. A set of post-hoc pairwise chi-square GOF tests were performed to investigate which one, if any, of the answers was significantly the most frequent, i.e., significantly more frequent than all other options. Because three comparisons were performed (option (1) against (2), (1) against (4), and (2) against (4)), a Bonferroni correction was applied, and the alpha-level was adjusted to $.05/3 = .017$ (the same procedure is followed in the rest of the post-hoc tests; where there are four potential answers to the questions, there are six comparisons and so the alpha-level is corrected accordingly to $.05/6 = .008$).

The comparisons showed, unsurprisingly, that *can in some cases stand alone as evidence* is significantly more frequent than both the other two options, with large effect sizes for both comparisons ($V = .78$ compared to *but can never stand alone*, and $V = .97$ compared to *other*). See the results of all chi-square GOF tests presented in sections 6.1 and 6.2, including the pairwise comparisons, in appendix B.

After each question, participants were able to leave an optional comment on the topic of that question as described in section 5.4. All these comments have been analysed (more on how below), but I will only include those analyses in this chapter in cases where the quantitative results called for extra interpretation (e.g., in the case of a high frequency of *other* answers, or in cases where there is a lack of a significant majority) or in cases where the qualitative analyses were judged to contain interesting patterns that were not available in the quantitative data. Overviews of the analyses of all the comments given by participants can be found in appendix C. The comments have not been edited in any way (not even to correct typos), though some have been shortened. This is marked with square brackets, "[...]".

[5] $V = 0$ signals no effect, and $V = 1$ signals a perfect effect. I will use Cohen's (1988) proposal for a conventional interpretation of effect sizes with $V = .1$ being a small effect, $V = .3$ being a medium effect, and $V = .5$ being a large effect.

The optional comments, as well as the alternative answers that participants who choose the *Supply other answer* option gave, were analysed according to the principles of content analysis (see Bauer 2000 for a review or Marvasti 2004, 90-94, for a brief introduction). Content analysis is a group of analysis methods ranging from impressionistic approaches on the one hand to structured and systematic methods on the other, used in different areas of social research (Rosengren, 1981). Marvasti (2004, 91) describes it as a method for "simplifying and reducing large amounts of data into organised segments" which are then used to interpret the values, norms, opinions, and attitudes of the community who produced the data. The general method is to go through the data material (which usually comes in the form of texts) and apply a set of codes which can either be developed bottom-up through the coding process or be pre-determined by the design and focus of the study. Then some form of quantification of the codes is performed (e.g., in the form of counts and descriptive statistics). Morgan (1993) argues that even though some sort of counting/quantifying is general to the method, one can distinguish a more quantitative use of the method from a more qualitative use. The quantitative approach is characterised by using pre-determined codes, which are applied automatically (perhaps by a piece of software). On the quantitative approach, the goal of the analysis is often, according to Morgan, to present the patterns that are present in the data. The qualitative approach to this method, on the other hand, more often uses codes that are developed bottom-up on the basis of the particular data-set. On this approach, the goal of the analysis is both to present the patterns found in the data and to interpret those patterns using the data.

Qualitative content analysis seems particularly apt for the analysis of the comments given in this study. On the one hand, it allows me to reduce large numbers of comments to a number of common themes and to present the found patterns in a succinct way. On the more qualitative side, it also provides a good supplement to the quantitative data from the questionnaire as it allows the codes to emerge from the comments. This is in contrast to the pre-specified options given in the questionnaire. It also allows me to use the comments to interpret what motivates the answers given by participants, which again is not possible from the quantitative answers themselves. In my presentation of the results of the content analyses, I follow Marvasti (2004). Comments were coded in an open, iterative coding-process to generate and systematise the codes. Two rounds of coding were carried out. This was done to systematise coding and to check consistency across coding sessions. Following the second round, some codes were updated for consistency (across questions) or organised into hierarchies of super- and subcategories of other codes. If a comment did not receive the same code in both rounds, both codes were considered, and a final

choice was made. All the coding was performed by me. In general, about 89% of codes were retained across coding rounds.

In the following paragraph, I show an example of a qualitative content analysis for the optional comments to question 1.

20 participants supplied optional comments on question 1. 7 of these comments (35% of the total number of comments to this question) are elaborations on the participant's view that intuitive judgements can stand alone as evidence. Another 4 participants (20%) mentioned that the details of the collection process are important. 3 participants (15%) mentioned that the answer to the question depends on a number of factors, e.g., which problem one is investigating. Another 3 participants (15%) challenged the view that intuitive judgements can stand alone as evidence, saying that they should only serve as the starting point of an investigation, or that in controversial cases it might be necessary to look at other types of evidence as well. The remaining comments mentioned certain irrelevant factors that might influence intuitive judgements (2, 10%) or contained a comment on the question (1, 5%). Table 6.2 presents an overview of the content analysis for these comments.

Theme	Example	No.	%
Can stand alone, elaborate	If they're no evidence, then we have virtually null information from all the grammar books based on intuitions. A highly unlikely idea.	7	35%
Focus on process	Context must be given with the test sentence	4	20%
Depends	It depends what kind of evidence you want - individual grammar or community info. [...]	3	15%
Not alone	... but should not be used alone.	3	15%
Other factors	Certain empirical phenomena is intrinsically connected to speakers judgments, e.g., ambiguity.	2	10%
Comment on question	It is not specified here _for what_ syntactic intuitions might constitute evidence.	1	5%
Total		20	100%

Table 6.2: Stand alone (Q1), optional comments, generative group

Summing up, the majority of participants in this group finds that intuitive judgements can stand alone as evidence. The comments corroborate the quantitative results but also allow for a more nuanced picture. The commenters stress that one should take care in the collection process, that there are certain situations in which relying on just intuitive judgements might not suffice, and that there are irrelevant effects that might influence intuitive judgements.

6.1.2 The etiology debate

In this subsection, I present the results of the analysis of the answers to the seven questions that were based on the main themes of the etiology debate (see chapter 2). For the order in which questions were presented in the questionnaire, see appendix A or section 5.4. All analyses are carried out as for question 1 above.

6.1.2.1 Origin – experience or competence

One of the main differences between the views in the etiology debate is why intuitive judgements can serve as evidence for grammatical theories. On the Voice of Competence view (VoC) as characterised by Devitt, as well as on the view presented by Devitt's critics, intuitive judgements can serve as evidence for grammatical theories because they are causally connected, in some way, to the speaker's linguistic competence. If either of these are the majority view among generative linguists, we would expect the *competence* answer (see below) to come out as most frequent. On the Modest Explanation, on the other hand, intuitive judgements can serve as evidence because they build on speakers' reflections on their empirical experience with language. If this view is the majority view, we would expect the *reflections* answer to come out as most frequent. On all three views, intuitive judgements are fairly immediate and unreflective, and the speaker does not have conscious introspective access to how the intuitive judgements are formed.

Participants were asked to complete the following sentence by choosing the option that expressed their opinion the best:

> **Question 2: Origin**
> When syntactic intuitions are reliable as evidence, this is *mainly* because ...
>
> 1. ... they are speakers' *reflections* about language use, and speakers are to some degree reliable judges about this.
> 2. ... they express speaker's *competence* in their native language.
> 3. [Supply other answer]

The results show that the generative majority view on this question is that intuitive judgements can serve as evidence because of their connection to speakers' competence. This supports VoC and the Critics' Views. The distribution of participants' answers is presented in table 6.3.

Answer	Reflections	Competence	Other
Frequency	17 (23%)	40 (55%)	16 (22%)

Table 6.3: Origin (Q2), distribution of answers, generative group

The answers were not equally distributed, $\chi^2(2, n = 73) = 14.30$, $p = .001$, $V = .31$. The pairwise comparisons showed that the *competence* answer is significantly more frequent than both *reflections* ($V = .40$) and *other* ($V = .43$).

Of the 16 participants who selected the *other* option, 7 (44%) wrote "both" or something to that effect. 3 participants (19%) answered that linguistic intuitive judgements make good evidence because they can tell us about competence.[6] For a full overview of the comments, see table 6.4. Note that one category of *other* answers in the table has been labelled "uncategorised". This will re-occur in a few places throughout the analyses. This code was used when the meaning of the comment was ambiguous or unclear to me.

Theme	Example	No.	%
Both	can be both	7	44%
Competence	they allow us to infer what the speaker's competence is	3	19%
Other	They reflect processing complexity, which could be labelled 'competence', but needn't be.	3	19%
Neither	Neither	1	6%
Comment on question	I do not understand the question	1	6%
Uncategorised	See below	1	6%
Total		16	100%

Table 6.4: Origin (Q2), other answers, generative group

Among the 10 participants who supplied an additional optional comment to this question, the comments were mixed. 3 participants (30%) supported the *competence* option, while another 3 participants (30%) pointed at other reasons than the provided options for linguistic intuitive judgements' status as evidence. See the full overview of the comments in table 6.5.

[6] As the *competence* option was already found to be significantly more frequent than the other answers, no new analysis was run with these three answers recoded.

Theme	Example	No.	%
Competence	[...] for sure it reflects something about the speaker's competence (defined in a very broad sense).	3	30%
Other	They are reliable because they are systematic. That is sufficient.	3	30%
Process	What we should take a given intuition to express depends at least in part on the task that was used to elicit the intuition.	2	20%
Not reflection	Option A is impossible: an intuition is never a reflection, it's just spontaneous and unreflected.	1	10%
Comment on question	[...] I find your simply binarity here misleading.	1	10%
Total		10	100%

Table 6.5: Origin (Q2), optional comments, generative group

The chi-square test gave a fairly straightforward result: There is a significant majority for intuitive judgements getting their status as evidence from their connection to competence. This, however, is not the whole story, as the distribution of the answers and the comments show. There is a non-negligible minority group supporting linguistic intuitive judgements getting their status as evidence from being reflections, and the comments show that some participants think that both sources might be relevant to linguistic intuitive judgements' status as evidence.

Overall, this result supports the claim that either VoC or the Critics' Views is the majority view among generative linguists.

6.1.2.2 Acceptability or grammaticality

A much debated question within the etiology debate is whether linguists should use intuitive judgements of *acceptability* or *grammaticality* as evidence for theories of grammar. On VoC and the Critics' Views, we should use intuitive judgements of acceptability as evidence in linguistics. On the Modest Explanation, on the other hand, the intuitive judgements we should use as evidence are judgements of grammaticality (see section 2.3.2). To introduce the acceptability/grammaticality distinction into the context, participants were first presented with the following text:

> "Most native speakers of English would find the sentence 'the dog

> that the woman that the man saw owned ran' to be an intuitively unnatural sentence in English. In the terminology of generative linguistics, the sentence is *unacceptable* to those speakers (this is not to be confused with whether or not the sentence is infelicitous to the speakers in some specific context). Most native speakers would also most likely find the sentence intuitively *ungrammatical*. However, some experts argue that it is, in fact, grammatical."

Participants were then asked to complete the following sentence by choosing the option that expressed their opinion the best:

> **Question 3: Acceptability/grammaticality**
> To the extent syntactic intuitions can serve as evidence for theories of grammar, only those syntactic intuitions can serve as evidence that are ...
>
> 1. ... acceptability intuitions.
> 2. ... grammaticality intuitions.
> 3. [Supply other answer]

There was no one significantly most frequent answer to this question. The distributions of the answers is presented in table 6.6. The chi-square GOF test showed that we cannot reject the null-hypothesis that the answers were equally distributed, $\chi^2(2, n = 74) = 4.08, p = .13$ (a similar result was found in the small generative group).

Answer	Acceptability	Grammaticality	Other
Frequency	31 (42%)	17 (23%)	25 (34%)

Table 6.6: Acceptability/grammaticality (Q3), distribution of answers, generative group

The quantitative result for this question seems inconclusive: We cannot say that either acceptability or grammaticality judgements are preferred as evidence for grammatical theories by a significant majority of participants. An interesting thing to look at here is the number of participants who chose the option *other*. Roughly one third of participants (25 out of 73) gave answers that were different from the two provided options. Of these 25, 15 participants (60%) answered "both" in some form or other (including two participants answering, essentially, "both, but mostly acceptability intuitions", and another two answering "both, but mostly grammaticality intuitions"). This suggests that a large number of participants were not happy with the stark contrast suggested in the question.

This is an interesting result, taking into consideration that the strong contrast that the question assumes is prominent in the generative literature. In the generative literature, we see claims that speakers do not have actual grammaticality intuitions, or that theoretical grammaticality judgements should not serve as evidence for grammatical descriptions in linguistics (see section 2.3.2). This is also part of VoC and the Critics' Views (at least in my understanding of those views). On the Modest Explanation, intuitive judgements can only be evidence for the grammar of a speaker's language if they are judgements of grammaticality (again, see section 2.3.2).

One might ask whether we could potentially re-code all the "both" comments as either *acceptability* or *grammaticality* depending on what we are interested in finding out. If, for instance, we want to know whether participants think that acceptability judgements *can* serve as evidence for theories of grammar, could we just add the 15 "both" answers to the 31 *acceptability* answers (total: 46) and re-run the tests? However, this changes the interpretation of the question that participants were asked, and so we cannot know for sure what their answers would have been, had the question been phrased that way.

Furthermore, as the question is phrased now, it sets up a mutually exclusive choice between acceptability judgements and grammaticality judgements, which, as we saw above, is based on the positions found in the literature. However, in answering "both", participants reject this mutually exclusive understanding of the situation. And so if we did add the "both" answers to either the *acceptability* answers (or the *grammaticality* answers), we could not use that result to conclude that, say, VoC is supported on this point, as a proportion of the participants whose answers contribute to this result *disagree* with VoC on this issue, specifically over whether *only* acceptability judgements can serve as evidence for grammatical theories. For these reasons, I do not re-code the comments in the suggested way.

The remaining 10 *other* answers were more mixed. 3 participants (12%) noted that the most important thing is how the linguistic intuitive judgements have been collected, e.g., that context and prosody should be controlled for. 2 participants (8%) answered that acceptability is the only real option.[7] Another 2 participants (8%) provided other suggestions for which type of linguistic intuitive judgements we should be collecting. A third group of 2 participants (8%) answered that they do not share the distinction of grammaticality vs. acceptability, again pointing to a problem with the stark contrast suggested by the question. The last participant (4%) answered that both types are problematic.

[7] One post hoc chi-square GOF test was re-run with these two comments recoded as *acceptability* rather than *other*. The relevant comparison (the one between *acceptability* and *grammaticality*) did not reach significance at the set alpha-level (with the relevant Bonferroni corrections for the number of comparisons, in this case $.05/3 = .017$).

See an overview of comments in table 6.7.

Theme	Example	No.	%
Both	Both, distinctly	15	60%
Focus on process	other factors must be controlled for carefully	3	12%
Acceptability	It is not possible to directly assess grammaticality intuitions; all judgements are acceptability judgements	2	8%
Other	felicity within a context	2	8%
Distinction	I do not share the distinction.	2	8%
Neither	I really see both as problematic.	1	4%
Total		25	100%

Table 6.7: Acceptability/grammaticality (Q3), other answers, generative group

One might worry that, since the terms "acceptability" and "grammaticality" are used somewhat interchangeably in the literature, participants would just choose the term that they most often use themselves without attending to the distinction as intended. The piece of text that was used to preface the question was added exactly to make the distinction salient in the context. All the participants whose answers are analysed in this section reported generative/formal linguistics as at least one of their influences, and so they can all be expected to be at least familiar with the distinction, even if some may use the terms interchangeably on a daily basis. The alternative answers provided by participants also show that at least within that group, some participants were attending to the distinction. Both the answers in the category "both", "acceptability", and "distinction" show awareness of and attention to the distinction, even if participants do not agree with the distinction or with the need to choose only one type of judgement.

23 out of 73 participants provided an optional comment on this question. 7 (30%) of these comments expressed the view that the answer could not possibly be *grammaticality*. This is, as mentioned in section 2.3.2, a common view in the literature. 4 participants (17%) commented on the difficulty in making a clear distinction between the two phenomena. 3 comments (13%) pointed to the importance of the process of how judgements are collected, and another 3 comments (13%) expressed the opinion that other types of linguistic intuitive judgements can serve as evidence as well. 2 participants (9%) elaborated on the link between grammaticality and acceptability, and 2 (9%) made direct comments on the question. A further 2 participants (9%) elaborated on their

view that grammaticality judgements are the ones that can serve as evidence. See an overview of the comments in table 6.8.

Theme	Example	No.	%
Not grammaticality	there is no such thing as a grammaticality intuition. grammaticality is a theoretical concept; acceptability is an empirical one.	7	30%
Distinction	It is not easy to draw a clear distinction between the two. [...]	4	17%
Focus on process	[...] This exx from English would need some context and appropriate prosody to accurately judged as grammatical.	3	13%
Other	The analyst can use all kinds of speaker judgments as evidence.	3	13%
Link	We can't access linguistic processes directly... in this regard, it's really difficult to tease apart 'acceptability' and 'grammaticality'	2	9%
Grammaticality, elaboration	But the theory will need to provide evidence-based arguments as to why such sentences are generally not accepted as well-formed by the speakers	2	9%
Comment on question	I don't think native speakers would find it 'ungrammatical' - I don't think they would use that term.	2	9%
Total		23	100%

Table 6.8: Acceptability/grammaticality (Q3), optional comments, generative group

The result of the test and the comments show a more nuanced picture than one might expect from the discussion in the literature. A large number of participants think that both acceptability and grammaticality judgements can serve as evidence (note that we did not ask whether they might play *some* role in research, such as spark new ideas, etc.; the question was specifically about the use of linguistic intuitive judgements *as evidence*). Aside from that, there are sizeable groups preferring both acceptability judgements, on the one side, and grammaticality judgements, on the other side.

The lack of a significant majority for either grammaticality or acceptability means that this result does not support any of the three main views of the

debate as the majority view among generative linguists. But furthermore, as discussed above, even if one were to add the participants who answered "both" to either side (those who think that intuitive judgements of acceptability can serve as evidence, or those who think that intuitive judgements of grammaticality can serve as evidence), we do not see clear support for any of the views of the debate. This is because it is part of those views that either only acceptability judgements can serve as evidence, or that only grammaticality judgements can serve as evidence. Thus, there is no majority support for VoC, the Critics' View, or the Modest Explanation on this issue among the generative participants in this study.

Finally, it should be noted that the lack of a significant result of the chi-square GOF test is potentially a problem related to the power of the test. There is a trade-off between power and the effect sizes a test can detect depending on the number of participants and the chosen alpha-level. With 73 participants, 2 degrees of freedom, the chosen alpha-level of .05, and the conventional power level of .8 (Cohen, 1988), effect sizes of .36 and above should be detectable (note that this number will differ according to the details of the test in question). While one should be careful with post-hoc interpretations of the power level of tests, this gives an indication that, if the test missed a true result, it would be at the lower end of the effect size scale. However, based on the literature, one would have expected to find a strong effect (in favour of acceptability), and at least we can conclude that this was not the case.

6.1.2.3 The role of the mental grammar

Another issue in the etiology debate is whether competence delivers the informational content of linguistic intuitive judgements or whether competence delivers data based on which the intuitive judgement is made. This is a central but rather abstract question. To get a handle on it, the question was phrased in terms of the only concrete proposal for a content-version of the story that has been proposed, namely the suggestion that linguistic intuitive judgements are "deduced" from the speaker's mental grammar (see section 2.3.3). If VoC is the majority view among generative linguists, we would expect the majority answer to be that intuitive judgements are likely deduced from the speaker's mental grammar. If either the Modest Explanation or the Critics' Views is the majority view, on the other hand, we would expect participants *not* to say that intuitive judgements are likely to be deduced from the speaker's mental grammar (i.e., to choose any other option).

Participants were asked to choose the option that expressed their opinion the best:

Question 4: The role of the mental grammar

Syntactic intuitions are sometimes said to be "deduced from the speaker's mental grammar".

1. This is probably a poor description of how intuitions are formed.
2. This is a good way to talk about how intuitions are formed but should probably not be taken too literally.
3. This is likely to be the actual process of how syntactic intuitions are formed in the mind.
4. [Supply other answer]

The *not too literally* option was significantly more frequent than the other options. See the distribution of answers in table 6.9.

Answer	Poor description	Not literal	Actual process	Other
Frequency	13 (18%)	44 (60%)	8 (11%)	8 (11%)

Table 6.9: Deduced (Q4), distribution of answers, generative group

The answers were not equally distributed, $\chi^2(3, n = 73) = 49.36, p < .001$, $V = .47$. The post-hoc analyses showed that the answer *not too literally* was significantly more frequent than *poor description* ($V = .54$) and both *actual process* and *other* ($V = .69$ for both comparisons).

This result shows that participants did not believe in the deduction-version of a story where competence provides the informational content of intuitive judgements. This does not support the claim that VoC is the majority view among generative linguists, but it is compatible with both the Modest Explanation and the Critics' Views being the majority generative view.

As mentioned, this question was intended to weigh on the question of whether competence supplies data or content for linguistic intuitive judgements. It is phrased in terms of deduction, as this is the only specific proposal that I know of for how the competence could provide the informational content of linguistic intuitive judgements with no input from the central processor. However, while no other alternatives are clearly formulated in the current stage of the debate, it is possible that a different version could be articulated. These results do not shed light on the potential support for such a suggestion. But, tentatively, I think we may assume that the majority view within generative linguistics is that the speaker's competence provides the data (not the content) for linguistic intuitive judgements.

6.1.2.4 Fallibility of linguistic intuitive judgements

A further point of discussion within the etiology debate is whether we have direct access to truths about our language through our linguistic intuitive judgements, not counting a certain level of noise from performance factors. On VoC we would expect participants to agree that we have this kind of access, while we would expect participants to disagree if the Modest Explanation is the majority view. The Critics' Views is unclear on this point. Participants were asked to indicate the degree to which they agree with the following statement:

> **Question 5: Fallibility**
> Imagine you could abstract away all performance factors that might influence a speaker's syntactic intuitions. In that case, it would be possible for the resulting syntactic intuitions to be mistaken about the grammatical properties of the sentence.
>
> 1. Strongly agree
> 2. Somewhat agree
> 3. Neither agree nor disagree
> 4. Somewhat disagree
> 5. Strongly disagree

Note that the categories *Strongly agree* and *Somewhat agree* were combined into one category (*agree*) for the analysis, as were the categories *Strongly disagree* and *Somewhat disagree* (*disagree*). The same goes for other questions in this and the following sections using this scale.

The majority view on this issue was to *disagree*. See the distribution of answers in table 6.10.

Answer	Agree	Neither	Disagree
Frequency	18 (25%)	12 (17%)	41 (58%)

Table 6.10: Fallibility (Q5), distribution of answers, generative group

The answers were not equally distributed, $\chi^2(2, n = 71) = 19.80, p < .001$, $V = .37$. Note that two participants explicitly asked to have their answers to this question disregarded, hence the lower number of participants in this test compared to other tests within the generative group. The post-hoc comparisons showed that *disagree* was significantly more frequent than both the *neither* option ($V = .55$) and the combined *agree* option ($V = .39$). Thus, the majority view on this issue is that if we could remove the performance factors from the equation, linguistic intuitive judgements would have to be correct about the

grammatical properties of sentences. On this view, a speaker has direct access to truths about their language trough their linguistic intuitive judgements, modulo some noise from performance factors. This result supports VoC being the majority view among generative linguists and does not support the Modest Explanation being the majority view. As the Critics' Views is unclear on this point, the result does not weigh either for or against this view being the majority generative view.

The optional comments highlighted some problems with the phrasing of this question. 20 participants in total supplied optional comments. 7 participants (35%) commented on the term "performance factors", and 4 (20%) expressed the opinion that the question was unclear (two of whom asked to have their answer disregarded, as mentioned above). Another 3 (15%) commented on the term "mistaken". These comments all pointed to a certain lack of clarity in the phrasing of the question. The rest elaborated their answers or had other comments on the question itself. See the full overview in table 6.11.

Theme	Example	No.	%
Performance factors	It very much depends on what is meant by 'performance factors'. Under some definitions, the answer is necessarily/definitionally 'yes'.	7	35%
Question unclear	This question is unclear to me, you should disregard my answer.	4	20%
Agree, elaborate	Linguistic illusions (work by Colin Phillips and many of his students) are a very good example of cases where speakers are 'mistaken' about the grammatical properties of sentences.	4	20%
Mistaken	[...] To be 'mistaken' is not a concept that I can meaningfully apply to intuitions, even if I imagine that performance is stripped off.	3	15%
Comment on question	This is a really interesting question!	2	10%
Total		20	100%

Table 6.11: Fallibility (Q5), optional comments, generative group

The comments on the phrasing of this question (14 in total out of 20) show that this question was unfortunately not as clear as intended. For instance, one

participant wrote: "I'm sorry, but this question is unclear to me. [...]" (and then goes on to comment on the term "performance factors"), and another wrote "I have difficulty understanding the statement. [...]" (and then goes on to comment on the term "mistaken"). See more examples in table 6.11. I take comments like those as a clear sign that participants were unsure about how to interpret the question. This means we should be careful in interpreting the quantitative results, even though they look clear at first glance.

The comments point to two ways in particular that the question may have been difficult to interpret. One is what is meant by "performance factors". The intended meaning of this term was the noise that comes from, e.g., memory limitations, lack of attention, etc., which are typically mentioned in the generative literature as influences that affect linguistic behaviour as well as acceptability judgements (Chomsky, 1965, 3). However, the term is not entirely specific, and so might be interpreted looser or stricter than on the traditional reading. If it is interpreted looser so as to include *all* irrelevant factors, then clearly linguistic intuitive judgements would be infallible if these factors could be filtered out. Such a view might seem to be obvious to someone who thinks that linguistic intuitive judgements are influenced only by the speaker's competence and by performance factors since on that view, all other factors than the speaker's competence are irrelevant. On the other hand, on a view where the central processor also plays a role, one could imagine a situation where, even with the performance factors filtered out, an intuitive judgement is mistaken about the grammatical properties of the sentence because of a faulty theory. The looser reading of the term on which "performance factors" covers all irrelevant factors, including ones that are not traditionally covered by the term, could lead to an inflation in the number of subjects who disagree with the statement.

On the other hand, there is also the (perhaps, slightly less likely) possibility of a stricter reading of the term on which only certain specific performance factors are included (say, memory limitations, but not attention span). This could have the opposite effect of the looser reading, leading some participants who do in fact think that intuitive judgements give direct access to truths about the speaker's language to agree with the statement because they find that certain sources of irrelevant noise are not covered by the term "performance factors". The stricter reading could thus lead to an inflation of the number of subjects who agree with the statement.

Another potential source of trouble is the phrasing that linguistic intuitive judgements could be "mistaken about grammatical properties of the sentence". This was intended to cover situations where the linguistic intuitive judgement is in conflict with the truth about the *grammaticality* status of the sentence for that speaker's language. However, on another reading, linguistic intuitive

judgements are correct if they correctly reflect the *acceptability* status of a particular sentence for a particular speaker. In this sense, they cannot be mistaken (unless the speaker for some reason does not accurately report how the sentence appeared to them). This reading would, I believe, lead to an inflation of the number of participants disagreeing with the statement.

Finally, the phrasing of the statement is somewhat convoluted, which might in itself have added to the difficulty of interpreting the question.

In conclusion, there are problems with the question pointing in different directions. There are two potential issues that might lead to an inflation of the number of participants who disagree with the statement, and the quantitative results did show a significant majority for the *disagree* option. Without further information, it is not possible to say whether the quantitative result would be similar to the one we see here if the statement was clarified to exclude the mentioned problems.

For that reason, we should only tentatively take this result as an indication of the real opinion of this group of participants. Taken at face value, the results show a majority view on which we would have access to truths about our own language trough our linguistic intuitive judgements if only we could weed out the influence from performance factors. This result provides (tentative) support for the claim that VoC is the majority view among generative linguists. It is not compatible with the Modest Explanation, and the Critics' Views is unclear on this point.

6.1.2.5 The subject matter of grammatical research

Participants were also asked to indicate their view on what the fundamental subject matter of grammatical research is. While this is not directly a question about linguistic intuitive judgements, different stances on this question tend to go with different opinions on linguistic intuitive judgements in the literature. If either VoC or the Critics' Views is the majority view, we would expect a majority of participants to say that they are ultimately interested in the mental linguistic capacity of the mind. If, on the other hand, the Modest Explanation is the majority view, we would expect a majority to say that they are ultimately interested in systematic patterns in linguistic behaviour. Participants were asked to complete the following sentence by choosing the option that expressed their opinion the best:

> **Question 6: Subject matter**
> When I study grammatical phenomena, I ultimately seek to understand ...
>
> 1. ... the systematic patterns found in linguistic behaviour.

2. ... the linguistic capacity of the mind.
3. [Supply other answer]

The distribution of answers can be found in table 6.12. There was no one significantly most frequent answer to this question within the generative group.

Answer	Patterns	Capacity	Other
Frequency	26 (36%)	33 (45%)	14 (19%)

Table 6.12: Subject matter (Q6), distribution of answers, generative group

The initial chi-square GOF test showed that the answers were not equally distributed, $\chi^2(2, n = 73) = 7.59, p = .02$, with a small effect size, $V = .23$. The post-hoc analysis, however, showed that no answer was significantly more frequent than both other answers. While the option *capacity* is significantly more frequent than the option *other* ($\chi^2(1, n = 47) = 7.68, p = .006$), it is not significantly more frequent than the answer *patterns* ($\chi^2(1, n = 59) = .83, p = .36$).[8] The results of the first test and the post-hoc comparisons are not necessarily contradictory. In general, it may be the case that the options are, overall, not equally distributed, while the two options with the most answers do not differ significantly in frequency (and that the difference in the first test is due to some of the other options). If we look at the distribution of answers presented in table 6.12, this looks like it could plausibly be the case here, as *capacity* and *patterns* are quite close in frequency while *capacity* and *other* are less alike in frequency.

Out of 14 participants providing other answers than the specified options, 13 participants (93%) answered, in one form or another, that they are interested in both systematic patterns and the linguistic capacity of the mind. Of these, 9 answered "both" in some form, 2 answered "both" plus something extra, and 2 answered that we are ultimately interested in the mental capacity but that we study it through the systematic patterns found in linguistic behaviour.[9] See the overview in table 6.13 (page 110).

As with the results for the Acceptability/grammaticality question (Q3, section 6.1.2.2), one might wonder whether we could potentially re-code all the "both" comments as either *patterns* or *capacity* depending on what we are interested in finding out. I will briefly recap the idea behind this thought. If, for instance, we want to know whether participants think that understanding "the linguistic capacity of the mind" is *a* valid goal of grammatical research

[8] As mentioned for the Acceptability/grammaticality question above (section 6.1.2.2), this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

[9] Re-coding these two answers as *capacity* rather than *other* does not lead to a significant result in the comparison between *capacity* and *patterns*.

Theme	Example	No.	%
Both	Both of the above	9	64%
Both and other	both, plus how these phenomena are acquired	2	14%
One through the other	The linguistic capacity of the mind by means of behavioural observations [...]	2	14%
Other	the processing mechanism(s) of language as they unfold in real time	1	7%
Total		14	100%

Table 6.13: Subject matter (Q6), other answers, generative group

(although perhaps not the ultimate goal), could we just add the "both" answers to the *capacity* answers and run the tests again? However, as mentioned above, this changes the interpretation of the question that participants were asked, and so we cannot know for sure what their answers would have been, had the question been phrased in this new way.

And like for the Acceptability/grammaticality question, the position expressed by participants writing "both" differs from the views found in the debate. As the question is phrased now, it asks about what researchers "*ultimately* seek to understand" (emphasis added) when studying grammatical phenomena. Assuming that participants answered the question with this phrasing in mind, the ones who chose the *other* option and wrote something like "both" reject both the view that the externally observable patterns in linguistic behaviour have priority as an object of study over the underlying mental linguistic capacities (which is part of the Modest Explanation), as well as the view that the mental linguistic capacities have priority as an object of study over patterns in linguistic behaviour (which is part of VoC and the Critics' Views). The position of these participants is thus one that is not represented in the debate.

So again, I do not re-code the comments in the suggested way, although I will note that the large number of participants who wrote "both" in some form must have an at least partially mentalist view on the aim of grammatical research.

Again in the optional comments, a sizeable number of participants (7 out of 15, 47%) answered that they are interested in both the mental capacities and the patterns of linguistic behaviour. See the full overview in table 6.14 (page 111).

Both the post-hoc tests and the comments point in the same direction here. Participants seem to think that both the mental capacities and the patterns of linguistic behaviour are fundamental to grammatical research. We did not see a majority subscribing to an interpretation grammatical research on which only the externally observable patterns are relevant (which is the case on the Modest

Theme	Example	No.	%
Both	I would have liked to answer 'both'	7	47%
Patterns and other	I want to find patterns and also learn something about the choices speakers make.	2	13%
Other, elaboration	Behaviour is not the end of the story, but we gather psycho-social info (unless doing implicit or neuro measures of language processing)	2	13%
Comment on question	Neither are the two given alternatives exhaustive, nor are they (mutually) exclusive.	2	13%
Capacity, elaboration	[...] I am working on a particular causal power which seems particularly relevant to language, which you have called 'the linguistic capacity of the mind'.	1	7%
Uncategorised	'linguistic capacity of the mind' is a bit vague for me.	1	7%
Total		15	100%

Table 6.14: Subject matter (Q6), optional comments, generative group

Explanation). On the other hand, we also did not see a majority subscribing to a view on which the ultimate reason to investigate the patterns found in language is to say something about the underlying mental capacities (as is the case on VoC and the Critics' Views). This is perhaps somewhat surprising given the strong mentalist commitment in the mainstream generative literature. The same result was found in the small generative group. This result does not support any of the three views from the debate being the majority view among generative linguists.

Still, as the 13 participants who answered "both" (in some form or another) have a conception of grammar that at least includes a mentalist component, one could argue that there is a majority for a conception of grammar that at least includes a mentalist perspective (46 participants in total, or 63% of the generative participants), although there is no majority for a purely mentalist conception of grammar.

6.1.2.6 Structure rules and mental rules

Another issue that comes up in the etiology debate, although it is not strictly about linguistic intuitive judgements in particular, is the question of how we should think about the structure rules we observe in linguistic data. On VoC, rules are implemented in the minds of speakers. This is not the case on the Modest Explanation. The Critics' Views is unclear on this point.

Participants were first introduced to the following text:

> "Syntactic intuitions is one type of evidence used by linguists to characterise how languages are structured. Let's call the rules that describe this the *structure rules* of particular languages."

Participants were then asked to choose the option that expressed their opinion the best:

> **Question 7: Rules**
> The structure rules that linguists describe are sometimes said to be "implemented in the minds of speakers".
> 1. It is a good hypothesis that structure rules are actually implemented in the minds of speakers.
> 2. From the structure rules we observe, we can only infer that the mind works *as if* it was following those rules.
> 3. From the structure rules we observe, we cannot infer anything about how the mind processes language.
> 4. [Supply other answer]

The majority view on this issue turned out to be that we can only infer that the mind works as if it was governed by structure rules. The distribution of participants' answers is presented in table 6.15.

Answer	Good hypothesis	As if	Nothing	Other
Frequency	21 (29%)	44 (60%)	5 (7%)	3 (4%)

Table 6.15: Rules (Q7), distribution of answers, generative group

The answers were not equally distributed, $\chi^2(3, n = 73) = 59.11, p < .001$, $V = .52$. Post-hoc comparisons showed that *we can only infer that the mind works as if it was following those rules* was significantly more frequent than the three other options, with effect sizes ranging from $V = .35$ to $V = .87$.[10]

A significant majority of generative participants thinks that we can only conclude that the mind works *as if* it was governed by the structure rules we find in the linguistic data. This result supports the Modest Explanation as the majority view among generative linguists, and it is not compatible with VoC being the majority view. As the Critics' Views was unclear on this point, the result weighs neither for nor against it being the majority generative view.

[10] In the small generative group, there was no view which was significantly more frequent than all others. In this group, 17 participants (30%) answered *good hypothesis*, 35 participants (61%) answered *as if*, 2 (4%) answered *nothing*, and 3 (5%) answered *other*. It is possible that the lack of a significant result in this case is due to the lower number of participants than in the large group. The distribution of answers is similar between the two groups.

6.1.2.7 The form of the implementation of mental rules

Finally, the participants who picked option (1) in question 7 above, i.e., who thought that it was a good hypothesis that structure rules are implemented in the minds of speakers, were then asked a follow-up question. Recall from section 2.3.6 that Devitt makes a distinction between a standard form of VoC on which rules are represented in the minds of speakers and a non-standard form on which rules are embodied in speakers' minds. If participants think that structure rules are implemented in speakers' minds, we would like to know if they think of those rules as being represented or embodied. Additionally, participants who answered *other* to the question above were also asked to answer this question (as one cannot anticipate what their answers might indicate). This means that 24 participants from the generative group were asked this question. Participants were asked to choose the option that expressed their opinion the best:

Question 8: Implementation

There must be something in the mind that gives rise to what we call "rules of grammar".

1. The rules of grammar are probably explicitly represented in the mind. If one could look into subjects' minds, one could find explicit rules.
2. The rules of grammar are implemented in the mind, but they are probably not explicitly represented.
3. The rules of grammar are probably not implemented in the mind.
4. [Supply other answer]

See the distribution of answers in table 6.16. The answer *not implemented* was included, partly as a safety hatch if one of the *other* answers from the previous question was not compatible with this question (in order to avoid forcing participants to be inconsistent in their answers) and partly to illustrate that this was not part of the intended reading of the *not represented* answer. As no participants chose this option, it was excluded from the analysis.

Answer	Represented	Not represented	Not implemented	Other
Frequency	12 (50%)	10 (42%)	0 (0%)	2 (8%)

Table 6.16: Implementation (Q8), distribution of answers, generative group

There was no one significantly most frequent answer to this question. The answers were not equally distributed, $\chi^2(2, n = 24) = 7, p = .03$, with a medium

effect size, $V = .38$. However, the two closest competitors, *represented* and *not represented* are fairly close in frequency as can be seen from table 6.16, and the post-hoc test, unsurprisingly, did not find that one answer was significantly more frequent than both other options. On this topic, then, the opinions seem to be divided.

To sum up, among the participants for whom this question was relevant, the opinions were divided, with roughly the same amount of participants thinking that the rules are likely represented and not represented. No answer was significantly more frequent than the rest, and so there was no majority for either the standard or the non-standard view among those participants who think that structure rules are implemented in the minds of speakers.

6.1.2.8 Summing up

In chapter 2, I presented the three main views of the etiology debate: the Modest Explanation, VoC, and the group of related views that I bundled together and called the Critics' Views. As mentioned in section 5.3, three formative indexes were created to cover the three views. Here, I review the results in light of these indexes and how the results play into the etiology debate.

Variable	Majority view	Modest Explanation	VoC	Critics' Views
Origin	Competence	0	1	1
Acceptability/ grammaticality	(No clear majority)	-	-	-
Deduced	Not too literally	1	0	1
Fallibility	Infallible	0	1	-
Subject matter	(No clear majority)	-	-	-
Rules	As if	1	0	-
Total		2	2	2

Table 6.17: Results: the three indexes, generative group

All the results are presented along with scores for each of the three indexes in table 6.17. As we can see from the table, no clear winner emerges from this comparison. The majority view that emerges from the analysis of the questionnaire answers agrees with each of the three views on exactly two points out of six. And, as noted in the analyses above, for some questions there simply is no clear majority view but rather a number of competing or co-existing views within the group.

This somewhat muddy picture is partly, I think, a reflection of the fact that there are more diverse opinions within this group than the somewhat polarised debate in the literature indicates. Another potential reason is that the indexes

are designed so that a tie could be the result (even if there had been clear majorities on all questions). This is of course a drawback of the design (as mentioned in section 5.3), but the design of the indexes was dictated by a close reading of the literature, and so variables could have only been cut or added if there were good reasons for doing so with the literature in mind. As mentioned, there might also be issues of phrasing for some questions, making some questions less clear than others, and finally a question of the balance between power and detectable effect sizes might mean that for the questions with no majority answer, weak effects might not have been detected because of the available number of participants (and general set-up). These points all urge us to only tentatively accept the results of the analysis.

Still, to sum up the majority view that emerges from this analysis: A significant majority of the generative participants think linguistic intuitive judgements can serve as evidence for grammatical theories because they reflect the linguistic competence of speakers. This result might be related to the following point: A significant majority of participants think that we would have infallible access to truths about our language through our linguistic intuitive judgements if we could filter out performance factors (though recall that several participants thought the phrasing of this question was unclear, meaning we should only tentatively count on this result being accurate).

The view that intuitive judgements would be infallible with performance factors filtered out does seem to make intuitive sense on a competence-based account of the etiology of linguistic intuitive judgements. If intuitive judgements can serve as evidence for grammatical theories because of their connection to competence (and not, e.g., because of us knowing true theories of linguistics), then if performance factors were filtered out we might expect the rest of the signal (from the competence) to be infallible.

On the question of whether we should think of structure rules as implemented in the minds of speakers or not, the majority view was that we can only say that the mind is structured *as if* those structure rules were implemented. Among those participants who found that it is likely that structure rules *are* implemented in the mind, there was no clear majority on whether we should think of those implemented structure rules as represented or not. When it came to the question of whether linguistic intuitive judgements are deduced from the rules that are implemented in the mind (whether those are the same as the structure rules we observe or not), the majority view was that we should not take such talk of deduction too literally.

Surprisingly, we also saw that opinions are divided on the issue of whether we should ultimately be interested in the abstract patterns found in linguistic data or the underlying mental capacities that give rise to them. The same

goes for the question of whether we should be using acceptability judgements or grammaticality judgements. For both questions, many participants answered "both" and thus expressed views that differ from the ones that are found in the debate (chapter 2).

It is perhaps surprising, taking the strong views expressed in the debate into consideration, that there was no clear majority on a number of these issues: whether we are ultimately interested in the abstract patterns in the linguistic data or the underlying mental capacity; whether we should use acceptability or grammaticality judgements as evidence; and, if structure rules are somehow implemented in the minds of speakers, whether they are represented or not. As mentioned above, the lack of significant results in these cases could be due to a problem with the power of the study, but even so it is at least still the case that the opinions among the participants of the study are not as clear-cut as the debate might lead one to expect.

The majority view that emerges within the generative group is competence-based as seen from the results on the questions about why linguistic intuitive judgements can serve as evidence and whether they would be fallible if performance factors were removed. This makes it differ from the Modest Explanation. However, the majority view is also not consistent with VoC as a majority of participants think that structure rules are not necessarily implemented in the minds of speakers and that intuitive judgements are not derived from the mental rules of speakers' grammars. The majority view does not correspond entirely to the Critics' Views either since it lacks significant majorities on some issues where the Critics' Views is clearly defined (acceptability judgements rather than grammaticality judgements, and mental capacities rather than external patterns).

In section 6.2, we will see that the non-generative group does not entirely match up with any of the three views from the debate either, but interestingly the majority view within that group is clearly closer to the Modest Explanation than to the other views.

6.1.3 Experimental Syntax

The questions presented in this subsection were included in the questionnaire to investigate what the current view among linguists is on the issues that are central to the Experimental Syntax debate (see chapter 3). Here, I look at the answers from the generative group. The analysis of these questions was performed as in the previous section.

6.1.3.1 Experts and ordinary speakers

The question of who to use as subjects when eliciting linguistic intuitive judgements is central in the Experimental Syntax debate. The experimentalists argue for the use of lay subjects, whereas the traditionalists argue that experts make for better subjects than non-experts. Participants were asked to complete the sentence below by choosing the option that expressed their opinion the best:

> **Question 9: Experts and ordinary speakers**
> The syntactic intuitions of linguists working on theories of grammar are, *on average*, ...
>
> 1. ... worse evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
> 2. ... equally good evidence for theories of grammar as the syntactic intuitions of ordinary speakers.
> 3. ... better evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
> 4. [Supply other answer]

There was no clear majority for any of the options. The distribution of the answers to this question is presented in table 6.18.

Answer	Better	Equally good	Worse	Other
Frequency	17 (23%)	28 (38%)	20 (27%)	8 (11%)

Table 6.18: Experts and ordinary speakers (Q9), distribution of answers, generative group

A chi-square GOF test showed that the four options were not equally distributed, $\chi^2(3, n = 73) = 11.2, p = .01$, $V = .23$ (small effect). However, the post-hoc pairwise comparisons of the options showed that no one option was significantly more frequently chosen than all of the other options (in fact, only the comparison between *equally good* and *other* was significant).[11,12]

As noted in table 6.18, eight participants chose the option *other* in response to question 9. Seven (87.5%) of these indicated that for them, the answer to this question depends on a number of factors, including the phenomenon under

[11] When running the same analysis for the small generative group, it was not possible to reject the null-hypothesis of equal distribution, $\chi^2(3, n = 57) = 7.1, p = .07$ (ns). This results in the same overall conclusion as for the full generative group: There is no clear majority for any of the options.

[12] As mentioned for the Acceptability/grammaticality question in section 6.1.2.2 above, this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

investigation, the area of grammar, the speaker, and elicitation method. An overview of the content analysis for these comments is given in table 6.19.

Theme	Example	No.	%
Depends	i think it depends on the speaker, the construction type, and how the judgments are elicited.	7	87.5%
Don't know	I don't know	1	12.5%
Total		8	100%

Table 6.19: Experts and ordinary speakers (Q9), other answers, generative group

Even though there is no significant majority for any of the options, it is interesting to note the proportions of answers each option receives. Since being a native speaker of a language should be enough to qualify a person to be a good source of linguistic intuitive judgements on the traditional view within generative linguistics, it is maybe surprising that roughly a quarter (23%) of participants in this group think that experts make better subjects than ordinary native speakers. On the other hand, the result is in line with the traditional practice of relying on the linguistic intuitive judgements of experts rather than those of ordinary speakers within generative linguistics. While this practice may be popular due to its convenience, the optional comments to this question suggest that there may be something more at play.

19 participants in total gave an optional comment on question 9. 5 participants (26%) made comments expressing the opinion that linguists make better subjects and qualifying why. Among the reasons they mention were that linguists are more likely to know what is being asked of them, that linguists are good at recognising (and avoiding) confounds, and that linguists' meta-linguistic knowledge leads to more nuanced judgements. Although the first of these reasons is convenience-based, the second two are less obviously about convenience. Instead, they suggest the view that linguists have a better access to the linguistic intuitive judgements that we all have as native speakers of our languages. While we do not know whether the commenters are representative for the 17 participants who choose the option *better*, their comments give us some insights into why this group might perceive linguists as being better subjects than naive speakers.

6 (32%) comments noted that it depends, there might be some benefits to using experts' linguistic intuitive judgements and other benefits to using the linguistic intuitive judgements of ordinary speakers. The most commonly mentioned benefits were more nuances on the side of the linguists, and being less

influenced by theoretical bias on the side of the ordinary subjects. 5 participants (26%) mentioned one way or another in which the linguistic intuitive judgements of linguists might be more subject to confounding factors than those of ordinary subjects, again with theoretical bias being the most commonly mentioned factor. An overview of the content analysis for all the optional comments on this question is given in table 6.20.

Theme	Example	No.	%
Depends	Linguists know to abstract away from things we think we are not interested in (like style, or frequency), but they may have theory-clouded judgments.	6	32%
Experts worse	It is known that repeated exposure to a marginally ill-formed structure weakens judgments.	5	26%
Experts better	Experts are generally good at recognizing confounds. That's what they're trained to do.	5	26%
Focus on process	The naive native speakers have to be given the right instructions, but they are probably easier to train in the correct methods. [...]	2	11%
Comment on question	The above response assumes, again, we're dealing with acceptability and not grammaticality judgments	1	5%
Total		19	100%

Table 6.20: Experts and ordinary speakers (Q9), optional comments, generative group

Summing up, this analysis shows that there are several groups with distinct views on this issue within the large generative group. This parallels the sides of the Experimental Syntax debate, with one side focusing on the benefits of using ordinary speakers and the other side focusing on the benefits of using expert subjects. Finally, the comments show that there is an awareness within the group about several factors that are deemed irrelevant to intuitive judgements which might nonetheless influence their content, such as theoretical bias, performance factors, and subjects not understanding the task properly. Some of these speak for the use of experts as informants while others speak for the use of ordinary speakers. As seen from the "depends" comments, several of the participants who gave optional comments drew the conclusion from these factors, which pull in different directions, that linguistic intuitive judgements from the

two different sources should be used under different circumstances.

6.1.3.2 Data and theoretical virtues

In the Experimental Syntax debate, the balance between reliable data and theoretical virtues is sometimes addressed. On the one side, some people argue that it is always most important to secure that the data one builds one's theory on is reliable. On the other side, it is argued that one may sometimes pay a little less attention to how the data is collected and instead focus on whether the theory one is building gives interesting and enlightening insights into the phenomenon under investigation (this position, which was also mentioned in section 3.4.3 and attributed to Grewendorf 2007, is sometimes referred to as Galilean linguistics, see, e.g., Chomsky 1980, and see Botha 1982 for a critical discussion). To investigate the distribution of these positions, we asked participants to indicate to what extent they agree with the following statement:

> **Question 10: Theoretical virtues**
> In general, if one has to choose, it is more important that a theory is built on reliable data than that it lives up to theoretical virtues such as simplicity, elegance, and fruitfulness.
>
> 1. Strongly agree
> 2. Somewhat agree
> 3. Neither agree nor disagree
> 4. Somewhat disagree
> 5. Strongly disagree

The results show that for the majority of this group, it is more important for a theory to be built on reliable data than for it to live up to theoretical virtues. The distribution of the answers can be found in table 6.21.

Answer	Agree	Neither	Disagree
Frequency	54 (74%)	14 (19%)	5 (7%)

Table 6.21: Theoretical virtues (Q10), distribution of answers, generative group

The options were not equally distributed, $\chi^2(2, n = 73) = 55.9, p < .001$, $V = .62$. Post-hoc pairwise comparisons show that *agree* is more frequent than both *neither* ($V = .59$) and *disagree* ($V = .83$).

Note that a minority of 7% disagreed with the statement, aligning themselves with the Galilean linguistic view on the matter. A further 19% neither agreed nor disagreed. Taken together, 26% of participants did not choose one of the

two *agree* options. As we will see below, a number of participants noted that both reliable data and theoretical virtues are important.

The optional comments suggest that this position is based, at least partly, on a wish to satisfy both conditions. There were 17 optional comments on this question. Of these, 7 comments (41%) pointed to the need for both. A further 7 (41%) were elaborations on why reliable data is more important than theoretical virtues, and 3 comments (18%) underscored the importance of theoretical virtues. An overview of the content analysis for the optional comments to question 10 is given in table 6.22.

Theme	Example	No.	%
Agree, elaborate	It doesn't matter how elegant a theory is if it accounts for the wrong data.	7	41%
Both	One always needs to balance compelling facts and compelling generalizations that don't capture all known facts. Welcome to science!	7	41%
Theoretical virtues important	[...] An elegant theory may contribute some truth for which we have yet to collect the right data in order to fully understand it. [...]	3	18%
Total		17	100%

Table 6.22: Theoretical virtues (Q10), optional comments, generative group

In conclusion, there is a large majority on the side of favouring reliable data over theoretical virtues. The group outright disagreeing with this stance is rather small. However, there are a number of voices that argue for a balance of both reliable data and theoretical virtues. While the need for reliable data is often used as an argument for Experimental Syntax, and while those who defend Galilean linguistics in the literature tend to also be on the traditionalist side of the Experimental Syntax debate, valuing reliable data over theoretical virtues is not a position that is exclusive to the experimentalist voices in the debate, and so this result cannot be taken as an indication that the majority of generative participants support Experimental Syntax.

6.1.3.3 Gradience in grammar

Another question that comes up within the Experimental Syntax debate is whether the gradience found in experimentally collected samples of intuitive judgements should be taken as evidence that the mental grammar is itself

graded, or whether the gradience is better interpreted as arising from interactions between grammar and other factors, e.g., working memory. In the survey, participants were asked to choose the one of the following statements that best expresses their opinion:

Question 11: Gradience

Some linguists use gradient rather than binary scales to collect syntactic intuitions.

1. Well-designed gradient scales may very well reflect real degrees of grammaticality.
2. Even well-designed gradient scales probably just capture effects that are not due to grammaticality.
3. [Supply other answer]

From the results, we cannot conclude that there is a clear majority view on this issue. Rather, there seems to be a number of competing views. The distribution of the answers to this question can be found in table 6.23.

Answer	Real gradience	Extra-grammatical	Other
Frequency	36 (49%)	22 (30%)	15 (21%)

Table 6.23: Gradience (Q11), distribution of answers, generative group

The answers were not equally distributed, $\chi^2(2, n = 73) = 9.4, p = .009$, $V = .25$. However, the post-hoc pairwise comparisons between the options showed that no option was significantly more frequent than both the two other options (and more specifically, the two closest competitors were not shown to be significantly different in frequency). According to these results, there is no clear majority generative view on this issue.[13]

Let us see what we might learn from the comments. 15 participants provided alternative answers to question 11. These answers were fairly mixed. Three participants (20%) commented that there is no gradience in grammar. Three participants (20%) commented that the gradience in experiments might result from both gradience in grammar and gradience in other factors.[14] A further two (13%) commented that they were not sure. An overview of the content analysis for the alternative answers to question 11 is given in table 6.24.

[13] As mentioned for the Acceptability/grammaticality question in section 6.1.2.2 above, this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

[14] As the participants answering "both" think that at least some of the gradience discovered is due to the nature of the grammar, these three answers could be recoded as *real gradience*. However, the three comments mentioned just above ("no gradience") could equally well be recoded as *extra-grammatical*. Recoding these six comments does not change the outcome of the test.

Theme	Example	No.	%
No gradience	There is no such thing as a 'degree of grammaticality'	3	20%
Both	A bit of both	3	20%
Don't know	Unsure	2	13%
Comment on question	difficult question :)	2	13%
Open question	I don't think we know yet whether gradience in judgments reflects real gradience or other effects. [...]	1	7%
Assumption	The assumption that everything sits on a single dimension of 'grammaticality' is likely a fiction.	1	7%
Uncategorised	They can certainly help.	3	20%
Total		15	100%

Table 6.24: Gradience (Q11), other answers, generative group

Eight participants gave optional comments to this question. Three (37.5%) noted that to them, this is still an open question. Two (25%) underscored their opinion that there is no gradience in grammar, and another two (25%) mentioned that the process of collecting and interpreting linguistic intuitive judgements is very important to the answer of this question. See the overview of comments in table 6.25.

Theme	Example	No.	%
Open question	still an open question; it is still not at all clear whether grammaticality is gradient or binary	3	37.5%
No gradience	Grammaticality IS binary. [...]	2	25%
Focus on process	They must be very 'well-designed'.	2	25%
Comment on question	It's very difficult to say what 'grammaticality' means in abstraction from a particular theory of grammar.	1	12.5%
Total		8	100%

Table 6.25: Gradience (Q11), optional comments, generative group

Within generative linguistics, there is a debate about gradience and about whether a gradient grammar is compatible with other traditional generative commitments (see chapter 3). On the one hand, one might argue that a gra-

dient grammar is hard to reconcile with the traditional goal of delineating the grammatical from the ungrammatical sentences of a language (though see Smith 1999, 32-33, for a suggestion that delineating sentences as "good" and "bad" was never part of the intended internalist, Chomskyan position, but see Carr 2003 for a critical discussion of this point). On the other hand, there are several proposals for how to integrate a graded grammar into the generative framework (see Schütze 2006/1996, ch. 3.3, and Sprouse 2013 for an overview). The divided result that arise for this question is possibly a reflection of this ongoing debate within generative linguistics.

To sum up, there are two competing views on this issue, and with the current set-up (number of participants, chosen alpha level etc.) no view was determined to be significantly more frequent than all others. It is possible, however, that a true result was missed due to lack of power of the test. The comments, however, corroborate the fragmented picture. There are both voices saying that the question is still open for debate, some saying that gradience might come from both the grammar and other factors, and some stating firmly that there is no gradience to be found in speakers' mental grammars.

6.1.3.4 Traditional methods

The main point of discussion within the Experimental Syntax debate is whether linguistic intuitive judgements should be collected and analysed through the informal, traditional armchair method or through formal, experimental methods. To investigate their opinion on this, participants were asked to indicate the degree to which they agree with the following statement (see also question 13 below):

> **Question 12: Traditional methods**
> In general, consulting one's own or one's colleague's syntactic intuitions produces good evidence for theories of grammar.
>
> 1. Strongly agree
> 2. Somewhat agree
> 3. Neither agree nor disagree
> 4. Somewhat disagree
> 5. Strongly disagree

The majority response to this question was agreement. See the distribution of answers in table 6.26 (page 125).

The answers were not equally distributed, $\chi^2(2, n = 73) = 37.0, p < .001$, $V = .50$. The post hoc pairwise comparisons showed that *agree* is significantly more frequent than both *disagree* ($V = .45$) and *neither* ($V = .75$).

Answer	Agree	Neither	Disagree
Frequency	48 (66%)	7 (10%)	18 (25%)

Table 6.26: Traditional methods (Q12), distribution of answers, generative group

20 participants provided optional comments on question 12. These comments were rather diverse. 3 (15%) participants mentioned that linguistic intuitive judgements collected this way can provide the starting point for an investigation. 3 (15%) found that linguistic intuitive judgements collected this way are okay in uncontroversial, stable cases, and another 3 (15%) said that it depends on the linguistic intuitive judgements being obtained in the right way. 2 (10%) mentioned that certain marginal analyses might be reinforced by in-group agreement among a number of experts when traditional methods are used. A full overview of the content analysis for these comments is given in table 6.27.

Theme	Example	No.	%
Starting point	Of course, this should only be the starting point.	3	15%
If uncontroversial	It usually produces robust and repeatable data, but there are often very subtle judgments that are not stable. [...]	3	15%
Focus on process	Assuming best practices are followed [...]	3	15%
Disagree, elaborate	In the long run, intuitions of individuals are just not enough. [...]	3	15%
Agree, elaborate	The proof is in decades of reliable findings [...]	2	10%
Group reinforcement	I see this leading to positive reinforcement of marginal analyses sometimes. Crony-ism	2	10%
Idiolects	Consulting one person is always good evidence of their own grammar [...]	1	5%
Many sources	It is good to have data points from as many sources as possible for control purposes.	1	5%
Grain of salt	but 'cum grano salis'	1	5%
Comment on question	Could I just comment that I think the phrasing 'evidence for theories grammar' is a bit off. [...]	1	5%
Total		20	100%

Table 6.27: Traditional methods (Q12), optional comments, generative group

6.1.3.5 Experimental methods

Similarly, participants were asked to indicate their level of agreement with the following statement about the use of experimental methods:

> **Question 13: Experimental methods**
> In general, syntactic intuitions should be collected and analysed by experimental methods from large numbers of speakers and using statistical tests.
>
> 1. Strongly agree
> 2. Somewhat agree
> 3. Neither agree nor disagree
> 4. Somewhat disagree
> 5. Strongly disagree

Again, the majority reaction to this question was agreement. See the distribution of answers in table 6.28.

Answer	Agree	Neither	Disagree
Frequency	45 (62%)	16 (22%)	12 (16%)

Table 6.28: Experimental methods (Q13), distribution of answers, generative group

The answers are not equally distributed, $\chi^2(2, n = 73) = 26.7, p < .001$, $V = .43$. The post hoc pairwise comparisons showed that *agree* is significantly more frequent than both *disagree* ($V = .58$) and *neither* ($V = .48$).

At a glance, the result for this question looks contradictory to the result for the previous one. In question 12, the majority of participants strongly or somewhat agreed that the traditional, informal method provides good linguistic evidence and in question 13, the majority of participants strongly or somewhat agree that linguistic intuitive judgements should be collected using experimental methods. However, the questions are not completely symmetrical and so are not mutually exclusive. Also, it is a possible opinion that both data types may have a role to play. The optional comments provided after these two questions shed some more light on the opinions of the participants.

26 participants provided optional comments on question 13. Out of these 26, 9 (35%) conveyed the opinion that while experimental methods are sometimes good and necessary, they are not always needed. 7 participants (27%) voiced the opinion that experimental methods are important for the cases where the judgements are controversial or very fine-grained. 3 participants (12%) mentioned that experimental methods can have a negative effect in the case of idiolects

(i.e., that one risks glossing over potentially meaningful variation in the data by averaging across a large number of speakers). Another 2 (8%) thought that experimental methods are useful when using them is feasible but stressed that this is not always the case, for instance in the case of endangered languages. A full overview of the content analysis for these comments is given in table 6.29.

Theme	Example	No.	%
Both	I tend to work this way, but it is not the only way and the different methods are mutually informative. [...]	9	35%
If controversial	If the data are contentious, then it may be necessary.	7	27%
Idiolects	Grammars vary and a small group of speakers could be enough.	3	11%
If feasible	if possible, yes, but in the case of endangered languages we work with what we have.	2	8%
Disagree, elaborate	I use those methods all the time. They rarely yield more than we get from simply asking a couple of friends.	1	4%
Agree, elaborate	[...] For two reasons: 1) this supplements the data, and more data is always better, and 2) it shows that the conclusions are representative. [...]	1	4%
Uncategorised	There is a literature about this.	3	11%
Total		26	100%

Table 6.29: Experimental methods (Q13), optional comments, generative group

Put together, the quantitative results and optional comments from question 12 and 13 show that there are many nuances to the generative participants' opinions on this issue. Overall, the participants find both the traditional, informal method and experimental methods useful. The traditional method may suffice in cases that are uncontroversial, serve as the starting point of an investigation, or be the best viable option in certain cases, like when working on endangered languages, or if one is interested in a particular person's idiolect. Experimental methods, on the other hand, may be preferable in cases where the judgements are controversial, unstable, or very fine-grained, or in cases where one suspects bias, such as group-reinforced agreement on marginal cases. These two results taken together do not provide clear-cut support for either the experimentalist

or the traditionalist side over the other.

6.1.3.6 Summing up

Overall, we did not see clear-cut support for one side of the Experimental Syntax debate over the other in the responses of the generative participants. The generative participants in this study do not agree on whether the linguistic intuitive judgements of experts provide better, worse, or equally good evidence for theories of grammar compared to the linguistic intuitive judgements of ordinary speakers. There is also no agreement on whether the gradience found in some experiments is due to grammatical or extra-grammatical factors. In both these cases, different opinions seem to exist along side each other. There is, however, general agreement that reliable data should be favoured over theoretical virtues when evaluating theories. Finally, the participants find both the traditional, informal method and experimental methods for collecting linguistic intuitive judgements useful. We saw through the comments that there is an awareness within the group about irrelevant factors that might influence the choice of source and method of collection for linguistic intuitive judgements.

6.1.4 Other questions and comments

In this subsection, I present the generative participants' answers to a couple of additional questions that do not fit into either the etiology or Experimental Syntax debates.

6.1.4.1 Significance

While methodological questions like the ones touched upon in this questionnaire are sometimes debated in linguistics, they do not take up much space, and the etiology question is rarely discussed. It would be interesting to know whether the issues touched upon in the questionnaire are important to linguists who use linguistic intuitive judgements as evidence or who believe that linguistic intuitive judgements may have that role. To answer this, participants were asked the following question:

> **Question 14: Significance 1**
> Can you think of a situation, before answering this survey, where you thought about whether syntactic intuitions can serve as evidence for theories of grammar?
>
> 1. Yes
> 2. No
> 3. [Supply other answer]

The distribution of the answers to this question is presented in table 6.30.

Answer	Yes	No
Frequency	70 (96%)	3 (4%)

Table 6.30: Significance 1 (Q14), distribution of answers, generative group

The answers were not equally distributed, $\chi^2(1, n = 73) = 61.5, p < .001$.[15] The effect size was very large with $V = .92$. An overwhelming majority of participants (96%) answered *Yes*.

11 participants gave optional comments on question 14. 7 of the comments (63%) give details of the circumstances under which participants have thought about the topic of the questionnaire. 3 comments (27%) mention the participant's reasons for giving thought to this question, including worries about extra-grammatical factors such as mathematical reasoning or prescriptivist attitudes interfering with linguistic intuitive judgements. An overview of the content analysis for the optional comments to question 14 is given in table 6.31.

Theme	Example	No.	%
When	in my own experiments on the topic :)	7	64%
Why	I have become worried that speakers can use extra grammatical reasoning, for instance mathematical reasoning, on things like quantifier scope.	3	27%
Significance	If anyone claims to be a linguist and answers 'no' to the above question then one must question whether they are, in fact, a linguist.	1	9%
Total		11	100%

Table 6.31: Significance 1 (Q14), optional comments, generative group

Participants were also asked to indicate their level of agreement with the following statement:

Question 15: Significance 2

Linguists who use intuitions as evidence should set aside time to consider *why* intuitions can serve as evidence for their theories.

1. Strongly agree

[15] 1 participant chose the *other* option and wrote "Yes" as their answer. This was recoded as the option *Yes*.

2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

The distribution of the answers to this question is presented in table 6.32.

Answer	Agree	Neither	Disagree
Frequency	65 (89%)	3 (4%)	5 (7%)

Table 6.32: Significance 2 (Q15), distribution of answers, generative group

For question 15, we also see a very large majority indicating that these issues are important by choosing *agree* (in total 89%). The answers are not equally distributed, $\chi^2(2, n = 73) = 102, p < .001$. The effect size, $V = .84$, is again very large. The post-hoc pairwise comparisons showed that *agree* is significantly more frequent than both *neither* ($V = .91$) and *disagree* ($V = .86$).

It is possible that the very high level of agreement among participants on these questions is a side-effect of the informal sampling method, where only people who find this topic interesting self-selected to participate in the questionnaire (and those who did not find it interesting did not). It is not possible to say whether this pattern would replicate with a randomly chosen sample of linguists, however it seems likely that such a strong effect would not be replicated.

12 participants gave optional comments on question 15. 9 (75%) out of these participants elaborated on why they answered *Strongly agree* or *Somewhat agree*, most of them by noting that reflecting on the data one uses is standard across scientific disciplines. An overview of the content analysis for the optional comments to question 15 is given in table 6.33.

Theme	Example	No.	%
Agree, elaborate	In the sense that all scientists should always take care and be reflective about methodology.	9	75%
Depends	It totally depends on the phenomenon one wants to analyze.	2	17%
Disagree, elaborate	Intuitions can tell us something about grammars, that's true. In my opinion, the question is what, not why. [...]	1	8%
Total		12	100%

Table 6.33: Significance 2 (Q15), optional comments, generative group

I take the participants' answers to these two questions (Q14 and Q15) to mean that the issues touched upon in the questionnaire are seen as significant by the participants. As mentioned, the overwhelming agreement on these questions might be a side-effect of the self-selection of participants: People who do already think the issue is important and worthwhile might be more likely to self-select to participate in the questionnaire. As such, the numbers reported here cannot be generalised to the linguist population at large. At least, though, they show that at least some proportion of linguists working with this type of data think that it is an interesting and important question why linguistic intuitive judgements can be used as evidence for grammatical theories.

6.1.4.2 Further comments

At the end of the questionnaire, participants were asked if they wanted to make any further comments on the topic of the questionnaire. 17 participants from the generative group did. 9 (53%) of these commented on the questionnaire, some pointing out things that could be improved, and others remarking that they found the topic interesting. Another 6 comments (35%) summed up a central part of the participant's view on linguistic intuitive judgements. The last 2 comments (12%) were about the importance of the topic to linguists. For an overview of these comments, see table 6.34.

Theme	Example	No.	%
Comment on questionnaire	I find that some of the given options suggest a binarity that just doesn't exist. [...]	9	53%
Central point	If native speakers' intuitive judgements cannot be used as evidence in linguistics, what at all can be?	6	35%
Significance	I think that linguists do not typically think about these issues	2	12%
Total		17	100%

Table 6.34: Further comments on the topic of the questionnaire, generative group

6.2 The non-generative majority view

In this section, I present the results from the group of participants who did not identify as formal/generative ($n = 61$). Of these, $n = 37$ identified as func-

tional/cognitive, $n = 12$ identified as having a mixed theoretical orientation (not including formal/generative influences), and $n = 12$ identified as atheoretical/other (see section 5.5). The analyses presented in this section were performed as for the generative group.

6.2.1 The etiology debate

In this subsection, I present the results of the analysis of the answers to the seven questions related to the etiology debate. In section 6.1.2, we saw that the majority view among the generative linguists was not VoC, but it also did not match entirely with the Modest Explanation or with the Critics' Views. In this section, we will see that the majority view on the etiology question among the non-generative participants in this study is closer to the Modest Explanation than to the two other views.

In short, on the majority non-generative view, the study of grammar is ultimately about the abstract patterns in language data rather than about the underlying mental linguistic capacity, the intuitive judgements we should use as evidence in linguistics are acceptability judgements, and intuitive judgements express speakers' reflections about language rather than their linguistic competence. On a further three questions, there was no significant majority for any of the options, but the way the answers pattern, along with the comments that participants provided, still allows us to say something about the group's view on these issues.

In the following, I take a look at each of the questions in turn and then sum up the view on the etiology question that emerges as the majority view among the non-generative participants.

6.2.1.1 Origin – experience or competence

Question 2: Origin (repeated)
When syntactic intuitions are reliable as evidence, this is *mainly* because ...

1. ... they are speakers' *reflections* about language use, and speakers are to some degree reliable judges about this.
2. ... they express speaker's *competence* in their native language.
3. [Supply other answer]

The answer *reflections* was significantly more frequent than the two other options. See the distribution of participants' answers in table 6.35.

Answer	Reflections	Competence	Other
Frequency	40 (66%)	14 (23%)	7 (11%)

Table 6.35: Origin (Q2), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(2, n = 61) = 29.74, p < .001$, $V = .49$. The post-hoc pairwise comparisons showed that the *reflections* answer was significantly more frequent than both *competence* ($V = .48$) and *other* ($V = .70$).

In other words, the non-generative participants in this study think that linguistic intuitive judgements make good evidence because they are based on speakers' reflections on language. This result is consistent with the Modest Explanation and in conflict with both VoC and the Critics' Views. It also contrasts with the majority view on this issue within the generative group, where the answer *competence* was significantly more frequent that the two other options.

6.2.1.2 Acceptability or grammaticality

Question 3: Acceptability/grammaticality (repeated)
To the extent syntactic intuitions can serve as evidence for theories of grammar, only those syntactic intuitions can serve as evidence that are ...

1. ... acceptability intuitions.
2. ... grammaticality intuitions.
3. [Supply other answer]

Acceptability was significantly more frequent than the other two options. The distribution of the answers is presented in table 6.36.

Answer	Acceptability	Grammaticality	Other
Frequency	36 (59%)	9 (15%)	16 (26%)

Table 6.36: Acceptability/grammaticality (Q3), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(2, n = 61) = 19.31, p < .001$, $V = .40$. The post-hoc comparison between *acceptability* and *grammaticality* had a large effect size of $V = .60$, and the comparison between *acceptability* and *other* had a medium effect size of $V = .38$. This result is consistent with VoC

and the Critics' Views and in contrast to the Modest Explanation. Within the generative group, there was no significant majority for any of the options.

6.2.1.3 The role of the mental grammar

Question 4: The role of the mental grammar (repeated)
Syntactic intuitions are sometimes said to be "deduced from the speaker's mental grammar".

1. This is probably a poor description of how intuitions are formed.
2. This is a good way to talk about how intuitions are formed but should probably not be taken too literally.
3. This is likely to be the actual process of how syntactic intuitions are formed in the mind.
4. [Supply other answer]

No one answer was significantly more frequent than all others. See the distribution of answers in table 6.37.

Answer	Poor description	Not literal	Actual process	Other
Frequency	28 (46%)	29 (48%)	1 (2%)	3 (5%)

Table 6.37: Deduced (Q4), distribution of answers, non-generative group

The initial chi-square GOF test showed that the answers were not equally distributed, $\chi^2(3, n = 61) = 46.21, p < .001$, $V = .50$. However, the post-hoc comparisons showed no one answer was significantly more frequent than all other answers.

When looking at the distribution of answers, one sees that the answer *poor description* and *not literal* are almost equally frequent with 28 and 29 responses respectively. The answer *actual process*, on the other hand, was only given by one participant. Even though there is no clear majority for any one answer, this pattern of answers is consistent with both the Modest Explanation and with the Critics' Views, though it is in contrast to VoC (on which intuitive judgements are hypothesised to be more or less directly deduced from the rules of the speaker's mental grammar). The pattern is also different from the one we saw among the generative participants where there was a significant majority for the option *not literal*.

Three participants provided different answers than the provided ones. See those comments in table 6.38.

Theme	Example	No.	%
Poor description	It's a poor description of how acceptability judgements are formed, but a good description of how intuitions about such things as what a sentence means are formed.	1	33%
Other	for me the constructionist approach seems more likely - when I reflect on my usage of L2 and L3, it looks like there really exist some chunks	1	33%
Uncategorised	This is a confused way of putting it	1	33%
Total		3	100%

Table 6.38: Deduced (Q4), other answers, non-generative group

There were five optional comments on this question. See the overview in table 6.39.

Theme	Example	No.	%
Not deduced	We might be able to deduce mental grammar from intuitions, but we cannot deduce intuitions from mental grammar, as we have access only to intuitions.	2	40%
Other	I would say that intuitions are due to the interiorization of a social rule	1	20%
Uncategorised	This should be true of authors of grammar books	1	20%
Comment on question	Don't think this question (or the options) are very clear. [...]	1	20%
Total		5	100%

Table 6.39: Deduced (Q4), optional comments, non-generative group

While there are too few comments to glean a pattern from, it is interesting that in both the *other* comments and in the optional comments, one comment suggests a very different etiology story for intuitive judgements than the one considered in the debate in generative linguistics. One comment suggests a constructionist approach, and another suggests that intuitive judgements are made based on internalised social rules. This is in line with the difference between the generative group and the non-generative group on this issue. In the generative group there was a significant majority for the option that deduction

talk may be a helpful way to talk about intuitive judgements although it should not be taken too literally. In the non-generative group, on the other hand, a sizeable number of participants thought deduction talk is just plainly a poor description of how intuitive judgements are formed.

6.2.1.4 Fallibility of linguistic intuitive judgements

Question 5: Fallibility (repeated)
Imagine you could abstract away all performance factors that might influence a speaker's syntactic intuitions. In that case, it would be possible for the resulting syntactic intuitions to be mistaken about the grammatical properties of the sentence.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

For this question, no one answer was significantly more frequent than the two others. See the distribution of answers in table 6.40.

Answer	Agree	Neither	Disagree
Frequency	29 (48%)	22 (36%)	10 (16%)

Table 6.40: Fallibility (Q5), distribution of answers, non-generative group

The chi-square GOF test showed that the answers were not equally distributed, $\chi^2(2, n = 61) = 9.08, p = .01$, with a small effect size of $V = .27$. However, on the basis of the pairwise post-hoc comparisons we cannot reject the null-hypothesis that the two closest competitors, *agree* and *neither* are equally frequent.[16]

This result contrasts with the result from the generative group, in which *disagree* was significantly more frequent than both other options. It is also in contrast with VoC, on which the preferred option would be *disagree*. On the Modest Explanation, the preferred option would be *agree*, so the pattern found in the non-generative group on this issue is partially in agreement with that view. It is unclear what the Critics' Views on this issue is.

Let us see if the comments can shed any further light on this situation. 18 participants supplied optional comments to this question. 4 participants (22%)

[16] As mentioned for the Acceptability/grammaticality question in section 6.1.2.2 above, this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

commented that performance factors are central to intuitive judgements. 3 participants (17%) used their comment to say that filtering out all performance factors is not possible. Both of these views could underlie the (relatively, compared to the generative group) low number of participants who chose *disagree*. Another 3 participants (17%) commented on an assumption of the question, and a further 3 participants (17%) made other comments on the question. 2 participants (11%) expressed the opinion that the distinction between competence and performance is questionable, and one participant (6%) wrote that the answer to the question depends on whether one is responding to clear cases or not. See the full overview in table 6.41.

Theme	Example	No.	%
Performance central	[...] Performance factors are part of the grammar.	4	22%
Not possible	I don't think it is ever possible to strip away performance factors	3	17%
Comment on question	Strangely formulated statement	3	17%
Assumption of question	This statement assumes that 'the' grammatical properties of the sentence exist completely separate from (and invariant among different) speakers	3	17%
Distinction	The distinction between 'competence' and 'performance' is a false distinction.	2	11%
Depends	If we are dealing with clear cases, I would say, 'I disagree'. If we are dealing with less-than-clear cases (by far the majority), I would say that I agree.	1	6%
Agree, elaborate	[...] Since patterns are found across sentences, not within one sentence, a speaker have lots of wrong intuitions about an isolated sentence.	1	6%
Uncategorised	Some people are not competent users of any language.	1	6%
Total		18	100%

Table 6.41: Fallibility (Q5), optional comments, non-generative group

The comments that performance factors are central to intuitive judgements is in contrast to the views put forward in the generative group (though recall that

there were some reservations about that result). It supports the general picture from the quantitative analysis that non-generative participants responded differently to this question than generative participants did. Significantly more participants in the non-generative group chose *agree* than *disagree*. Together with the comments, this signals that, on the non-generative view, intuitive judgements do not give us direct access to a mental grammar which can, at least in principle, be separated from performance factors.

6.2.1.5 The subject matter of grammatical research

Question 6: Subject matter (repeated)
When I study grammatical phenomena, I ultimately seek to understand ...

1. ... the systematic patterns found in linguistic behaviour.
2. ... the linguistic capacity of the mind.
3. [Supply other answer]

The distribution of answers can be found in table 6.42. The answer *patterns* was significantly more frequent than both other options.

Answer	Patterns	Capacity	Other
Frequency	42 (69%)	10 (16%)	9 (15%)

Table 6.42: Subject matter (Q6), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(2, n = 61) = 34.66, p < .001$, $V = .53$. The answer *patterns* was significantly more frequent than both *capacity* ($V = .62$) and *other* ($V = .65$). This view is in line with the Modest Explanation, but in contrast to both VoC and the Critics' Views. It also contrasts with the majority view within the generative group, where there was no significant majority for any of the options.

In sum, there was a clear majority among the non-generative participants for the view that we are ultimately interested in studying the patterns in linguistic data rather than the mental underpinnings of those patterns.

6.2.1.6 Structure rules and mental rules

Question 7: Rules (repeated)
The structure rules that linguists describe are sometimes said to be "implemented in the minds of speakers".

1. It is a good hypothesis that structure rules are actually implemented in the minds of speakers.

2. From the structure rules we observe, we can only infer that the mind works *as if* it was following those rules.
3. From the structure rules we observe, we cannot infer anything about how the mind processes language.
4. [Supply other answer]

No answer to this question was significantly more frequent than the three other options. The distribution of the answers is presented in table 6.43.

Answer	Good hypothesis	As if	Nothing	Other
Frequency	5 (8%)	34 (56%)	17 (28%)	5 (8%)

Table 6.43: Rules (Q7), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(3, n = 61) = 37.03, p < .001$, $V = .45$. The post-hoc pairwise comparisons, however, showed that no one answer was significantly more frequent than each of the three other options. More specifically, the option *as if* was not significantly more frequent than the option *nothing*, its closest competitor.[17]

Although this result does not give us a clear majority view on this issue, it can still help us understand something about what *is not* the non-generative majority view on this issue. It is clearly not the case that the majority thinks that intuitive judgements are "deduced" from speakers' mental grammars (the *good hypothesis* answer). In this sense, this result is in contrast to VoC. On the Modest Explanation, the processing rules of the mind must respect the structure rules that we observe (the *as if* option), and while the result does not show that this option is significantly more frequent than all other options, it is significantly more frequent than the *good hypothesis* option, which is arguably the most interesting competitor (*as if* is also significantly more frequent than the *other* option). In the generative group, *as if* was significantly more frequent than all other options, including *nothing*. This suggests that while a section of the two groups agree on the *as if* option, there is a difference between the two groups in that a larger proportion of the non-generative linguists chose the option that structure rules tell us *nothing* about how the mind processes language (in section 6.3.2.2, we see that there is in fact a significant association between participants' theoretical orientation and their answers to this question).

The 5 alternative answers (*other*) are along the same lines as the quantitative results. Of those 5 participants, 2 indicated that claims such as the ones made

[17] As mentioned for the Acceptability/grammaticality question in section 6.1.2.2 above, this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

in the options should only be viewed as hypotheses. One participant answered that, although we can only infer that the mind works as if the structure rules are implemented, it is nonetheless a good hypothesis that they are implemented.[18] Another participant commented that it might altogether be wrong to talk about "rules" in this context, and the final participant expressed the view that structure rules result from observing language use. See an overview of the comments in table 6.44.

Theme	Example	No.	%
Just hypothesis	[...] In principle, this hypothesis is acceptable as such, but to test it, psycholinguistic and brain research would be needed.	2	40%
Both	Both A and B. We can only infer that the mind works 'as if' it was following those rules — but it is nonetheless a good working hypothesis that the rules are in fact implemented in the minds of speakers.	1	20%
Not rules	Rules may not be the best conceptualisation of the process	1	20%
Other	I'd say structure rules result from observing language; they can help monitoring language production when the speaker needs help	1	20%
Total		5	100%

Table 6.44: Rules (Q7), other answers, non-generative group

11 participants provided optional comments to this question. 4 of these comments were elaborations on the answer *nothing*. 2 were comments on the question, and 2 were uncategorised. 1 participant elaborated on the *good hypothesis* answer, 1 participant expressed the opinion that claims such as the ones in the options are just hypotheses, and one participant noted that the answer depends on what is meant by "how the mind processes language". See an overview of the optional comments in table 6.45 (page 141).

The comments show a similar pattern to the quantitative results. Several comments stress how this is a hypothesis that should be tested, which shows a degree of reservation towards the *good hypothesis* answer. Another chunk of comments focus on how the underlying cognitive reality might be very different from the structure rules that we can observe. As with the distribution of the

[18]Re-coding this answer as *as if* rather than *other* does not lead to a significant result.

Theme	Example	No.	%
Nothing, elaborate	[...] Correlating rules that linguists arrive at with cognitive reality is haphazardous.	4	36%
Comment on question	Poorly worded question. [...]	2	18%
Uncategorised	If forced to choose between the three options, that is the one I would choose.	2	18%
Good hyp., elaborate	Of course, a good hypothesis can still be falsified	1	9%
Just hypothesis	From the rules alone we can only infer the 'as if'. Based on that, we hypothesize an implementation. [...]	1	9%
Depends	The answer also depends on the definition of 'how the mind processes language'.[...]	1	9%
Total		11	100%

Table 6.45: Rules (Q7), optional comments, non-generative group

quantitative answers, this suggests that the non-generative group of participants tend more towards the *as if* and *nothing* answers than towards immediately accepting that it is a good hypothesis that structure rules are implemented in the minds of speakers.

6.2.1.7 The form of the implementation of mental rules

Question 8: Implementation (repeated)
There must be something in the mind that gives rise to what we call "rules of grammar".

1. The rules of grammar are probably explicitly represented in the mind. If one could look into subjects' minds, one could find explicit rules.
2. The rules of grammar are implemented in the mind, but they are probably not explicitly represented.
3. The rules of grammar are probably not implemented in the mind.
4. [Supply other answer]

As mentioned in section 6.1.2.7, the answer *not implemented* was included, partly as a safety hatch if one of the *other* answers from the previous question was not compatible with the other options of this question (not forcing participants to be inconsistent in their answers) and partly to illustrate that this was not part of the intended reading of the *not represented* answer. As no participants in this group chose this option, it was excluded from the analysis.

10 participants from the non-generative group answered this question (see section 6.1.2.7). See the distribution of answers in table 6.46.

Answer	Represented	Not represented	Not implemented	Other
Frequency	2 (20%)	4 (40%)	0 (0%)	4 (40%)

Table 6.46: Implementation (Q8), distribution of answers, non-generative group

There was no one significantly most frequent answer to this question. The chi-square GOF test showed that we cannot reject the null-hypothesis of equal distribution, $\chi^2(2, n = 10) = 0.8, p = 0.67$.

Out of the 10 participants who answered this question, 4 participants chose to supply their own answer, and they each had different perspectives. One participant commented that it might not be right to talk of "rules" in this case. Another participant commented that the connection between the mind and observable patterns might not be transparent. One participant commented that the rules of grammar *may* be implemented in the mind (my emphasis). Finally, the last participant commented on the question itself. See all four comments in table 6.47.

Theme	Example	No.	%
Not rules	Rules may not be the best conceptualisation of this process.	1	25%
Not transparent	There must be an interface between whatever is in the mind andobservable patterns of grammar. But that doesn't mean the mind operates on what we observe. [...]	1	25%
Other	The rules of grammar may be implemented in the mind.	1	25%
Comment on question	Is the mind the same as the brain?	1	25%
Total		4	100%

Table 6.47: Implementation (Q8), other answers, non-generative group

3 participants provided optional comments on this question. Of these, 2 were elaborations on the *not represented* answer and one was an elaboration on the *represented* answer. See the comments in table 6.48.

Theme	Example	No.	%
Not repr., elaborate	I am doubtful that much of anything is 'explicitly represented in the mind', beyond perhaps direct sensory data. [...]	2	67%
Repr., elaborate	[...] Firstly it, unlike the two alternatives, instantly explains how come there seem to be rules. Secondly it explains why language is so simple and regular ('below the surface').	1	33%
Total		3	100%

Table 6.48: Implementation (Q8), optional comments, non-generative group

Among the participants for who this question was relevant, the opinions were divided, with neither *represented* or *not represented* being significantly more frequent than the rest. There were only a few comments, but they mainly expressed reservations of one sort or another towards the idea that structure rules are represented in the minds of speakers.

6.2.1.8 Summing up

Like in section 6.1.2.8, I here review how the majority answers from the non-generative group fit the three main views of the etiology debate (as measured by the three formative indexes presented in chapter 5).

Variable	Majority view	Modest Explanation	VoC	Critics' Views
Origin	Reflections	1	0	0
Acceptability/ grammaticality	Acceptability	0	1	1
Deduced	(No clear majority)	1	0	1
Fallibility	(No clear majority)	-	0	-
Subject matter	Patterns	1	0	0
Rules	(No clear majority)	-	0	-
Total		3	1	2

Table 6.49: Results: the three indexes, non-generative group

The results are presented along with scores for each of the three indexes in table 6.49. As we can see from the table, the non-generative group of participants agree more with the Modest Explanation than with either of the two other views. The main difference is that, on the majority non-generative view, we should use intuitive judgements of acceptability rather than of grammatical-

ity as evidence for grammatical theories, which is in contrast with the Modest Explanation.

On the two other issues where the majority non-generative view did not align exactly with the Modest Explanation, it was the case that two options were tied in frequency. In both these cases, one of the two positions that were tied agree with the Modest Explanation. In the question about structure rules and mental rules, one of the two tied positions is that the mind just works "as if" it was governed by the structure rules that we observe. This is in line with Devitt's "respect" position (see section 2.3.6). In the question about whether linguistic intuitive judgements are infallible once performance factors are filtered out, one of the tied positions is that intuitive judgements, even in this case, would be fallible, which is in line with the Modest Explanation as well.

Overall, the majority view within the non-generative group is quite close to the Modest Explanation with the exception that it is speakers' judgements of acceptability, rather than their judgements of grammaticality, that should figure as evidence in linguistics. This marks a clear difference from the majority generative view, which incorporated elements from each of the three views in the debate. That the non-generative view is close to the Modest Explanation is perhaps not surprising. Devitt argues that, although on his account intuitive judgements may serve as evidence, it will often be better to rely on linguistic usage data. This view is widespread in non-generative linguistics as well (see section 4).

To sum up, on the non-generative majority view that emerges from this analysis: Linguistic intuitive judgements can serve as evidence for theories of grammar because they express speakers' reflections about linguistic behaviour, the judgements we should use as evidence are judgements of acceptability rather than grammaticality, and the ultimate goal of grammatical theory is to study the externally observable patterns in linguistic behaviour. As for the remaining three issues where there was no significant majority for any of the options, we can still say something about what the majority view is *not*: It is not part of the majority non-generative view that structure rules are themselves implemented in the minds of speakers. Rather, from the structure rules we observe we can either only say that the mind works as if it followed those rules or they tell us nothing at all about the structure of the mind. It is also not part of the majority non-generative view that intuitive judgements are deduced from the rules of speakers' mental grammars. Finally, it is not part of this view that linguistic intuitive judgements would give us direct, infallible access to truths about speakers' languages if only we were able to filter out performance factors.

6.2.2 Experimental Syntax

In this subsection, I present the non-generative answers to the questions that are based on the Experimental Syntax debate (see chapter 3).

In short, the results are the following: Among the non-generative participants, a significant majority favoured the position that experimental methods should in general be used when collecting and analysing linguistic intuitive judgements. There was also a majority in favour of the position that it is generally more important that a theory is built on reliable data than whether it lives up to theoretical virtues. Finally, there was a significant majority for the position that the gradience found in some judgement experiments is likely due to real degrees of grammaticality. On the issues of whether the traditional method of collecting intuitive judgements provides good evidence for grammatical theories and whether linguists' intuitive judgements are better evidence for grammatical theories than those of ordinary speakers, there were no significant majorities. Overall, this suggests that the non-generative group falls closer to the experimentalist side of the Experimental Syntax debate than to the traditionalist side.

6.2.2.1 Experts and ordinary speakers

Question 9: Experts and ordinary speakers (repeated)
The syntactic intuitions of linguists working on theories of grammar are, *on average*, ...

1. ... worse evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
2. ... equally good evidence for theories of grammar as the syntactic intuitions of ordinary speakers.
3. ... better evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
4. [Supply other answer]

There was no clear majority for any of the options. The distribution of the answers is presented in table 6.50.

Answer	Better	Equally good	Worse	Other
Frequency	10 (16%)	21 (34%)	25 (41%)	5 (8%)

Table 6.50: Experts and ordinary speakers (Q9), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(3, n = 61) = 17.10, p < .001$, $V = .31$. However, the post-hoc pairwise comparisons showed that no option was significantly more frequently chosen than all of the other options.[19] When looking at the distribution of answers, the two options *equally good* and *worse* are clearly close to each other in frequency. This result is similar to the one observed in the generative group, although in the generative group it looks like slightly more participants chose *better* and slightly fewer chose *worse* than in the non-generative group (in the association analysis in section 6.3.2, we see that this difference is not significant).

The optional comments provided by participants for this question can shed a bit more light on the participants' reasons for choosing the particular answers they selected. 20 participants in the non-generative group gave an optional comment on question 9. 10 (50%) of these comments expressed the idea that there are more confounds involved when using experts' intuitive judgements than when using those of lay subjects. They mention that linguists might carry more uncontrolled, unconscious biases such as confirmation bias and bias from their formal knowledge of linguistic theory. Some also mention that the professional experience of a linguist might change their sense of what structures are acceptable, for instance through exposure to data from other languages. Four participants (20%) noted that it depends on, for instance, characteristics of the participants, e.g., whether the linguists are in general better educated than the lay subjects, and whether the linguist is prescriptively or descriptively minded. Other participants noted that it depends on how one understands "better": Linguists might on the one hand understand the task more easily, but they might on the other hand bring more theoretical bias to the task, or linguists might be more open to accepting marginal structures than lay subjects, but they might also more easily be misled by structures which are theoretically "possible" but which a native speaker would not recognise. 2 participants (10%) elaborated on their answer that experts make better subjects than lay speakers. They write that linguists are more accurate than lay subjects, and that linguists understand the reasoning behind the question and are better equipped to analyse the structures they are presented with. An overview of the content analysis for all the comments to this question is given in table 6.51 (page 147).

The distribution of the quantitative results and the optional comments given by participants show that opinions are divided on this issue within the non-generative group of participants. From the comments, we see that many participants worry about the potential confounds which might influence the intuitive

[19] As mentioned for the Acceptability/grammaticality question in section 6.1.2.2 above, this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

Theme	Example	No.	%
More confounds	The linguist will be biased by his formal knowledge or beliefs regarding grammar...	10	50%
Depends	assuming their level of education is higher than that of an average person.	4	20%
Better, elaborate	[...] linguists are more accurate about languages than native speakers.	2	10%
Uncategorised	I am unsure about the theretical (and practical) relevance of the concept of native speaker.	2	10%
Not alone	...but reliance on introspection alone is not good scientific practice	1	5%
Different	Not worse, not better, not equal, just different, reflecting different experience and exposure.	1	5%
Total		20	100%

Table 6.51: Experts and ordinary speakers (Q9), optional comments, non-generative group

judgements of linguists, but we also see some voices arguing that there might be certain benefits to using linguists as subjects. This result is, perhaps surprisingly, quite similar to the one obtained from the generative group.

6.2.2.2 Data and theoretical virtues

Question 10: Theoretical virtues (repeated)
In general, if one has to choose, it is more important that a theory is built on reliable data than that it lives up to theoretical virtues such as simplicity, elegance, and fruitfulness.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

In the non-generative group, there was a very large majority who thought that it is more important for a theory to be built on reliable data than for it to live up to theoretical virtues. The distribution of the answers to question 10 can be found in table 6.52.

Answer	Agree	Neither	Disagree
Frequency	55 (90%)	3 (5%)	3 (5%)

Table 6.52: Theoretical virtues (Q10), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(2, n = 61) = 88.66, p < .001$, $V = .85$. The post-hoc pairwise comparisons show that *agree* is more frequent than both *neither* and *disagree* ($V = .90$ for both comparisons). A minority of 5% disagreed with the statement, and a further 5% neither agreed nor disagreed. Taken together, only 10% of participants did not choose one of the two *agree* options.

In conclusion, there is a large majority on the side of favouring reliable data over theoretical virtues among the non-generative participants, and both the group that outright disagrees with this stance and the one who neither agrees nor disagrees are rather small. This result is similar to the one observed in the generative group.

6.2.2.3 Gradience in grammar

Question 11: Gradience (repeated)
Some linguists use gradient rather than binary scales to collect syntactic intuitions.

1. Well-designed gradient scales may very well reflect real degrees of grammaticality.
2. Even well-designed gradient scales probably just capture effects that are not due to grammaticality.
3. [Supply other answer]

In the non-generative group, the answer *real degrees* was significantly more frequent than the other two options. See the distribution of the answers in table 6.53.

Answer	Real gradience	Extra-grammatical	Other
Frequency	41 (67%)	13 (21%)	7 (11%)

Table 6.53: Gradience (Q11), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(2, n = 61) = 32.39, p < .001$, $V = .52$. The pairwise comparisons showed that *real degrees* was significantly more frequent than both *extra-grammatical* ($V = .52$) and *other* ($V = .71$). This result differs from the one found in the generative group, where there was no significant majority for any of the three options.

In other words, there was a significant majority for the view that gradient scales express real degrees of grammaticality within the group of non-generative participants. While one might worry about how gradience fits with the traditional generative view of grammar (see the discussion of the results for this

question for the generative group in section 6.1.3.3), there is no general commitment to grammar being binary in non-generative (functional/cognitive) theories of grammar. In other words, the *real gradience* answer does not immediately clash with any particular commitments in this group.

6.2.2.4 Traditional methods

Question 12: Traditional methods (repeated)
In general, consulting one's own or one's colleague's syntactic intuitions produces good evidence for theories of grammar.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

No answer was significantly more frequent than both the other options for this question. See the distribution of answers in table 6.54.

Answer	Agree	Neither	Disagree
Frequency	17 (28%)	12 (20%)	32 (52%)

Table 6.54: Traditional methods (Q12), distribution of answers, non-generative group

The answers were not equally distributed, $\chi^2(2, n = 61) = 10.66, p = .005$, $V = .30$. The post hoc pairwise comparisons showed that none of the three options were significantly more frequent than both other options. More specifically, the option *disagree* was not found to be significantly more frequent than the option *agree*, and therefore we cannot reject the null-hypothesis of equal distribution.[20]

This result differs from what we saw in the generative group, where the answer *agree* was found to be significantly more frequent than the two other options. Looking at the optional comments that participants gave, we may get some insight into some of the reasons for this difference.

15 participants from the non-generative group provided optional comments on this question. 7 (47%) participants thought that informally collected linguistic intuitive judgements can serve as a good starting point of investigations but that they should be followed up by other methods. 4 (27%) thought that

[20] As mentioned for the Acceptability/grammaticality question in section 6.1.2.2 above, this lack of a significant result could potentially be due to a problem with the power of the test for the current set-up.

it depends on factors such as whether we are dealing with clear cases, who the linguist is (and who their colleagues are), and on the domain of interest. 3 (20%) participants elaborated on their reason for choosing *disagree*. A full overview of the content analysis for these comments is given in table 6.55.

Theme	Example	No.	%
Starting point	these intuitions can be a good starting point to do some serious corpus-derived and experimental work	7	47%
Depends	again, it is ok if we are dealing with clear cases. Otherwise, I would look for empirical data	4	27%
Disagree, elaborate	I do not accept intuitions as of much validity.	3	20%
Uncategorised	cf. my comment to the last question	1	7%
Total		15	100%

Table 6.55: Traditional methods (Q12), optional comments, non-generative group

A large number of these comments express reservations in some form or another about using informally collected intuitive judgements. This reflects the lower degree of participants agreeing that informally collected intuitive judgements can function as evidence for theories of grammar found in the non-generative group compared to the generative group.

6.2.2.5 Experimental methods

Question 13: Experimental methods (repeated)
In general, syntactic intuitions should be collected and analysed by experimental methods from large numbers of speakers and using statistical tests.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

A significant majority of participants in the non-generative group answered *agree* to this question. The distribution of answers is found in table 6.56.

Answer	Agree	Neither	Disagree
Frequency	46 (75%)	10 (16%)	5 (8%)

Table 6.56: Experimental methods (Q13), distribution of answers, non-generative group

The answers are not equally distributed, $\chi^2(2, n = 61) = 49.21, p < .001$, $V = .64$. The post hoc pairwise comparisons showed that *agree* was significantly more frequent than both *neither* ($V = .64$) and *disagree* ($V = .80$). We saw a similar result within the generative group, although the generative group also had a significant majority that agreed that the traditional method produces good evidence (Q12).

Summing up, a significant majority of non-generative participants agree that, in general, formal methods should be used to collect and analyse linguistic intuitive judgements.

6.2.2.6 Summing up

The non-generative group of participants in this study do not agree on whether the linguistic intuitive judgements of experts provide better, worse, or equally good evidence for theories of grammar compared to the linguistic intuitive judgements of ordinary speakers. There is, however, a majority agreement that we should prioritise reliable data over theoretical virtues and that the gradience found in some intuitive judgement experiments is really due to gradience in the grammar. Finally, there was a significant majority of the non-generative participants who agreed that, in general, formal methods should be used to collect and analyse intuitive judgements. There was no significant majority on the issue of whether informally collected intuitive judgements make good evidence or not. Based on this set of results, the non-generative group of participants agree more with the experimentalist side of the Experimental Syntax debate than with the traditionalist side. Within the generative group, we saw that opinions were more divided among the two sides of the debate.

6.2.3 Other questions and comments

In this subsection, I present the analysis of the non-generative participants' answers to those questions that are not related to either the etiology or the Experimental Syntax debates specifically.

6.2.3.1 Can syntactic intuitions stand alone as evidence?

Question 1: Stand alone (repeated)
When studying grammatical phenomena, syntactic intuitions ...

1. ... can in some cases stand alone as evidence.
2. ... can in some cases serve as evidence but can never stand alone.
3. ... can in no way be used as evidence.
4. [Supply other answer]

As mentioned earlier, this question served as one of the inclusion criteria for this study. Only participants who thought that intuitive judgements can in some way serve as evidence were included in the study (and the analysis of this question).

Within the non-generative group, there was no significant majority on this question. See the distribution of answer in table 6.57. From the table, it is clear that *in some cases* and *never alone* are very close together in frequency.

Answer	In some cases	Never alone	Other
Frequency	27 (44%)	33 (54%)	1 (2%)

Table 6.57: Stand alone (Q1), distribution of answers, non-generative group

The initial chi-square GOF test showed that the answers were not equally distributed, $\chi^2(2, n = 61) = 28.46, p < .001$, $V = .48$. However, the pairwise comparisons did not find one answer to be significantly more frequent than both other options.

In the generative group, the answer *in some cases* was significantly more frequent than both other options. This difference is perhaps not surprising, as intuitive judgements have a less prominent place in non-generative linguistics than it does within generative linguistics. Throughout the results of this questionnaire, we also see in several cases that the non-generative group has more reservations about the use of intuitive judgements as evidence for grammatical theories than the generative group does.

6.2.3.2 Significance

Question 14: Significance 1 (repeated)
Can you think of a situation, before answering this survey, where you thought about whether syntactic intuitions can serve as evidence for theories of grammar?

1. Yes
2. No
3. [Supply other answer]

The distribution of answers to this question is presented in table 6.58. The answers were not equally distributed, $\chi^2(1, n = 61) = 39.36, p < .001$, with a very large effect size, $V = .80$. An overwhelming majority of participants (90%) answered *Yes*.

Answer	Yes	No
Frequency	55 (90%)	6 (10%)

Table 6.58: Significance 1 (Q14), distribution of answers, non-generative group

Participants were also asked to indicate their level of agreement with the following statement:

Question 15: Significance 2 (repeated)
Linguists who use intuitions as evidence should set aside time to consider *why* intuitions can serve as evidence for their theories.

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

See the distribution of answers in table 6.59. For question 15, we also see a very large majority indicating that these issues are important by choosing either *Strongly agree* or *Somewhat agree* (in total 89%).

Answer	Agree	Neither	Disagree
Frequency	54 (89%)	4 (7%)	3 (5%)

Table 6.59: Significance 2 (Q15), distribution of answers, non-generative group

The answers are not equally distributed, $\chi^2(2, n = 61) = 83.64, p < .001$, $V = .83$. The post-hoc pairwise comparisons showed that *agree* is significantly more frequent than both *neither* ($V = .86$) and *disagree* ($V = .89$).

These results mirror the ones obtained in the generative group (in the association analyses described in section 6.3, no significant differences between the theoretical groups were found for these questions), and like for the generative group we see very large effect sizes in the tests. As discussed in section

6.1.4.1, this might be an issue of the sampling in which participants self-selected (and so, researchers who were already interested might have been more likely to participate). Since the use of linguistic intuitive judgements as evidence for grammatical theories is arguably more widespread in generative linguistics than in non-generative linguistics, one would maybe have expected that more generative linguists than non-generative linguistics would indicate that they have previously thought about the issue (question 14) . However, as only participants who thought that linguistic intuitive judgements can serve as evidence in this way were included in the study, the result is perhaps less surprising.

To sum up, these results show that the issues touched upon in the questionnaire are seen as significant by the participants, even though the large effect sizes perhaps suggest a bias in the sampling. However, the results at least show that some of the linguists working with this type of data think it is an interesting and important question why linguistic intuitive judgements can be used as evidence.

6.2.3.3 Further comments

At the end of the questionnaire, participants were asked if they wanted to make any further comments on the topic of the questionnaire. 16 participants from the non-generative group did. 10 (62.5%) of these commented on the questionnaire, some pointing out things that could be improved, and others remarking that they found the topic interesting. Another 4 (25%) participants commented on their preferences when it comes to data use. Three of those preferred using corpus data, and one commented that they preferred not using intuitive judgements. The last 2 comments (12.5%) expressed an additional point about topics of the questionnaire which the participant found important. For an overview of these comments, see table 6.60.

Theme	Example	No.	%
Comment on questionnaire	An interesting survey!	9	56%
Data preference	I prefer corpus data (which entail a lot of methodological and theoretical problems as well). Syntactic intuitions is sometimes a useful complement.	4	25%
Additional issues	I reject the importance of being a 'native speaker'.	3	19%
Total		16	100%

Table 6.60: Further comments on the topic of the questionnaire, non-generative group

6.3 Association analysis

In addition to the analyses presented in the previous sections, for each individual question it was also analysed whether specific background variables were associated with different answers to that question. Throughout this section I will note which analyses were done, but only the significant results are reported. The first analysis below is described in detail as an example of how all the analyses in this section were carried out.

For each question, potential associations between answers and three different background variables were tested:

- Theoretical group: within the group of all participants, i.e., between the large generative group and the non-generative group, and within the large generative group, i.e., between those who exclusively identified as formal/generative and those who identified formal/generative linguistics as just one of their influences (see section 5.5).
- Specialisation within linguistics (see the full list in appendix A).
- Stance on methodology: experimentalist vs. traditionalist (as determined based on answers to questions 12 and 13, see explanation in section 6.3.1).

An association analysis can answer questions about whether a relationship exists between two variables (in this case, answers to a question in the questionnaire and, say, theoretical orientation). The null-hypothesis that is tested in each case is that there is no association between the two variables in question. It is similar to a correlation analysis, which is done with interval data (and sometimes ordinal data). Association analyses can be done with nominal data as well. I use Fisher's exact test of association (also called Fisher's exact test of independence).[21]

6.3.1 Example analysis

Recall question 1, the question asked at the opening of the questionnaire. As in section 6.1.2, I go through the analysis of the answers to this question in considerable detail to illustrate how the analysis was done for all the questions in this section. Again, it may be skipped or skimmed if one is familiar with the method of analysis used, or one may return to it later as a reference. Recall that only participants who thought that intuitive judgements can in some way serve as evidence were included in the study (and the analysis of this question).

[21] Fisher's exact test was used instead of the similar chi-square test as the expected frequency for each cell of the contingency tables for the analyses in this section is low, which violates an assumption of the chi-square test.

Question 1: Stand alone (repeated)
When studying grammatical phenomena, syntactic intuitions ...

1. ... can in some cases stand alone as evidence.
2. ... can in some cases serve as evidence but can never stand alone.
3. ... can in no way be used as evidence.
4. [Supply other answer]

I first present the analysis of whether there is an association between answers to question 1 and theoretical orientation, then for whether answers to this question and linguistic specialisation are associated, and finally for whether answers to this question are associated with participants' stance on methodology.

6.3.1.1 Theoretical orientation

Fisher's exact test of association was used to test whether there was an association between the theoretical orientation of participants and their answers to question 1. The test was non-significant within the generative group ($p = .50$). That means that there was no significant difference in answers to this question between those participants who identified solely as formal/generative and those who chose the option *mixed* and mentioned formal/generative linguistics as one of their influences (see section 5.5). For the formal/generative group vs. the non-generative group, however, the test was significant, $p < .001$, with a medium effect size of $V = .47$. The distribution of answers is presented in table 6.61.

Answer / Group	Generative	Non-generative
Can stand alone	64 (88%)	27 (44%)
std. res.	5.36	-5.36
Can never stand alone	8 (11%)	33 (54%)
std. res.	-5.39	5.39
Other	1 (1%)	1 (2%)
std. res.	-0.13	0.13
Total	73 (100%)	61 (100%)

Table 6.61: Association analysis: Stand alone (Q1) and theoretical orientation, all participants

The numbers in the table below the row with frequency count and percentage are the standardised residuals. The significant result in Fisher's exact test just tells us that there was *some* significant difference in how the

two groups answered the question, it does not tell us what that difference consists of. Standardised residuals can help tell us this, as they are a measure of how significant each particular cell in the table is to the overall difference between the groups. The standardised residual for a particular cell is $(observed\ count - expected\ count)/\sqrt{expected\ count}$.

The standardised residuals are calculated with a chi-square test. Based on Sheskin (2003, 226-227), standardised residuals larger than ± 1.96 are counted as significant at the .05 level in two-tailed tests. To see whether an answer is significantly more or less frequent for one group compared to the other, one inspects the standardised residual for that cell, and if the number is larger than 1.96 then this group is significantly more likely than the other group to choose this answer. If the number is lower than -1.96, this group is significantly less likely to choose this answer than the other group. If the number is somewhere in between, the answers to this questions do not contribute to the overall significance of the association.

The standardised residuals in the table show that the participants in the generative group were more likely to find that linguistic intuitive judgements can stand alone as evidence than the non-generative group, with participants in the non-generative group being more likely to find that linguistic intuitive judgements cannot stand alone as evidence than the generative group.

One thing to note is that group A being more likely than group B to give answer 1 does not mean that answer 1 is the most frequent answer in group A. To see that this is the case, imagine that both groups might have answered 2 most frequently, but whereas in group B *all participants* answered 2, only 80% of participants in group A chose 2, while 20% chose 1 – in that case, group A would be more likely than group B to answer 1. We shall see examples of this later on. All it signifies is that participants from group A are more likely to give this particular answer than participants from group B are, no matter the overall frequency of the answer within each of the groups.

6.3.1.2 Specialisation

In the background part of the questionnaire, participants were asked to choose 1-3 areas of specialisation from a list of 16 possible specialisations within linguistics (see section 5.5 or appendix A). They were also given the option to specify another specialisation if they found that their specialisation was missing. To test whether area of specialisation is associated with the answers to question 1, a test was carried out for each of the 16 options with one group being all participants who had indicated this particular specialisation and the other group being all

other participants.[22] The test was performed both a) within the group of all participants, b) within the large generative group, c) within the small generative group, and d) within the non-generative group. This gives 64 combinations to test for each question. Only the significant results are reported.

For question 1, the only significant associations with specialisations were within the group of all participants for the following specialisations: syntax ($p = .004, V = .27$), sociolinguistics ($p = .008, V = .28$), and corpus linguistics ($p = .002, V = .28$). The effect sizes of these associations are all around small to medium size. In table 6.62, the distributions of answers for these three groups are presented together. There were no significant associations between answers to question 1 and specialisations within any of the other groups.

Answer / Specialisation	Not syntax	Syntax	Not socio	Socio	Not corpus	Corpus
Can stand alone	31 (55%)	60 (77%)	85 (73%)	6 (35%)	75 (76%)	16 (46%)
std. res.	-2.64	2.64	3.08	-3.08	3.27	-3.27
Can never stand alone	25 (45%)	16 (21%)	30 (26%)	11 (65%)	23 (23%)	18 (51%)
std. res.	2.99	-2.99	-3.27	3.27	-3.11	3.11
Other	0 (0%)	2 (3%)	2 (2%)	0 (0%)	1 (1%)	1 (3%)
std. res.	-1.21	1.21	0.54	-0.54	-0.77	0.77
Total	56 (100%)	78 (100%)	117 (100%)	17 (100%)	99 (100%)	35 (100%)

Table 6.62: Association analysis: Stand alone (Q1) and specialisations, all participants

From this analysis, we see that within the group of all participants, syntacticians were more likely to think linguistic intuitive judgements can stand alone as evidence than non-syntacticians. Sociolinguists and corpus linguists are both less likely to think this than the groups of non-sociolinguists and non-corpus linguists respectively, and both groups were more likely than their counterparts to think that linguistic intuitive judgements cannot stand alone as evidence.

6.3.1.3 Traditionalists vs. experimentalists

Finally, it was analysed whether answers to question 1 were associated with participants' stance towards experimental and traditional ways of collecting and analysing linguistic intuitive judgements. For this purpose, participants' scores on the two questions about methods, question 12 and 13, were turned into numerical values according to table 6.63 (page 159).

For each participant, the scores for the two question were added together. This results in five possible scores: $-2, -1, 0, 1, 2$. On this scale, a negative

[22] As participants were able to specify more than one specialisation, it was not possible to perform one test on the contingency table for the whole set of specialisations since the observations would not be independent of each other (one participant would appear in multiple columns), violating an assumption of the test.

Answer(s)	Strongly agree, Somewhat agree	Neither agree nor disagree	Somewhat disagree, Strongly disagree
Score Q12	1	0	-1
Score Q13	-1	0	1

Table 6.63: Scores for the traditionalist-experimentalist scale

score indicates agreement with the experimentalist side of the debate, while a positive score indicates agreement with the traditionalist side of the debate. These scores were then tested for association with the answers to question 1. The test was done for the group of all participants, the large generative group, the small generative group, and the non-generative group.

There was a significant association between scores on the traditionalist-experimentalist scale and answers to question 1, but only for the group of all participants ($p < .001, V = .31$). The distribution of answers can be found in table 6.64.

Answer / Group	-2	-1	0	1	2
Can stand alone	20 (51%)	12 (52%)	30 (71%)	17 (94%)	12 (100%)
std. res.	-2.64	-1.78	0.59	2.59	2.50
Can never stand alone	19 (49%)	10 (43%)	12 (29%)	0 (0%)	0 (0%)
std. res.	2.92	1.47	-0.34	-3.03	-2.41
Other	0 (0%)	1 (4%)	0 (0%)	1 (6%)	0 (0%)
std. res.	-0.91	1.24	-0.96	1.53	-0.45
Total	39 (100%)	23 (100%)	42 (100%)	18 (100%)	12 (100%)

Table 6.64: Association analysis: Stand alone (Q1) and stance on methods, all participants

From table 6.64 and the standardised residuals, we see that those who agree more with the traditionalists (positive scores) are more likely to think that linguistic intuitive judgements can stand alone as evidence than those who do not. On the other hand, those who have a score of -2, and so agree more with the experimentalists, are more likely to think that linguistic intuitive judgements cannot stand alone (the trend is similar for those with a score of -1, but this result does not reach significance).

Now, I will present the results for the rest of the questions. First, I present the results of the analyses of association with theoretical orientation, then specialisation, and finally stance on the methodology issue. In the following, I do not present tables like the ones above for each question. Instead, I present the conclusions from one or more questions together in overview tables. For tables like the ones above for all the analyses in this section, see appendix D.

6.3.2 Theoretical orientation

First, let us take a look at potential differences between participants who identified solely as formal/generative and those who identified as "mixed" and mentioned formal/generative linguistics as one of their influences (see section 5.5).

There was only one significant association between answers to the questions of the questionnaire and theoretical orientation within the generative group. This suggests that there are no large differences in opinions on the questions of the questionnaire between those who identify solely as formal/generative (the small generative group) and those who see formal/generative linguistics as one of their influences (the mixed group). In sections 6.1.2 and 6.1.3, we saw the same pattern: In almost all cases, we found the same significant results in the large and small generative groups.

The significant association was with answers to question 5 (Fallibility; $p = .04, V = .32$). The mixed group (n = 16) was significantly more likely to say *neither agree nor disagree* than the small group (n = 57). The mixed group was somewhat less likely than the small group to disagree that linguistic intuitive judgements could be mistaken if all performance factors were abstracted away from, but this difference did not reach significance (standardised residual: -1.57). This result is presented in table 6.65. Note that the small generative group was not significantly more likely to pick any of the other categories than the mixed group. This is marked by a dash in the table.

Question	p	V	Small	Mixed
Fallibility (Q5)	.041	.32	-	Neither

Table 6.65: Overview of associations with theoretical orientation, within the generative group

In the rest of this section, I look at the group of all participants to see if answers to particular questions were associated with participants' categorisation into either the generative group or the non-generative group.

There were several significant associations between answers to the questions of the questionnaire and whether participants belonged to the formal/generative group or not. I will present this part of the analysis question by question below. See a summary of the results in table 6.66. Note again that these results do not necessarily identify the most frequent answer for the groups but answers where the groups differed significantly from each other.

Question	p	V	Generative	Non-generative
Origin (Q2)	$< .001$	.43	Competence	Reflections
Deduced (Q4)	.001	.34	Actual process	Poor description
Fallibility (Q5)	$< .001$	.42	Disagree	Agree, Neither
Subject matter (Q6)	$< .001$	.35	Capacity	Patterns
Rules (Q7)	$< .001$	.36	Good hypothesis	Cannot infer anything
Theoretical virtues (Q10)	.032	.22	Neither	Agree
Traditional methods (Q12)	$< .001$	.38	Agree	Disagree

Table 6.66: Overview: associations with theoretical orientation, all participants

6.3.2.1 Origin – experience or competence

For this question, there was a significant association between answers and theoretical group ($p < .001, V = .43$). The non-generativists were more likely than generativists to find that linguistic intuitive judgements are good evidence because they are speakers' reflections about language, whereas the generativists were more likely to think that linguistic intuitive judgements are good evidence because of their relation to competence than non-generativists (the distribution of answers for each group to this and the remaining questions can be found in the relevant tables in section 6.1.2 to 6.2.2 and will not be repeated here).

6.3.2.2 The role of the mental grammar

There was a significant association between answers to this question and theoretical orientation ($p = .001, V = .34$). The non-generativists were more likely, compared to generativists, to answer that the deduction story is a poor description of how linguistic intuitive judgements are formed. Generativists, on the other hand, were more likely than non-generativists to answer that linguistic intuitive judgements might actually be deduced from the mental grammar (neither of these positions are the majority positions in the respective groups).

6.3.2.3 Fallibility of linguistic intuitive judgements

There was a significant association between theoretical orientation and answers to this question ($p < .001, V = .42$). Non-generativists were more likely than generativists to agree that linguistic intuitive judgements can be mistaken even if performance factors are abstracted away from, and they were also more likely to answer *neither agree nor disagree*. Generativists, on the other hand, were more likely than the non-generativists to disagree with the statement that linguistic intuitive judgements could be mistaken.

6.3.2.4 The subject matter of grammatical research

The association analysis showed a significant association between answers to question 6 and theoretical orientation ($p < .001, V = .35$). The generative group was more likely to think that we should ultimately be interested in the mental linguistic capacity than the non-generative group, who in turn was more likely to think that we should ultimately be interested in the abstract linguistic patterns than the generative group.

6.3.2.5 Structure rules and mental rules

The analysis showed a significant association between theoretical orientation and answers to this question ($p < .001, V = .36$). The generative group was more likely to answer that it is a good hypothesis that structure rules are implemented in the minds of speakers than the non-generativists (note that this is not the most frequent position among generativists). The non-generative group was more likely to answer that we cannot infer anything about how the mind processes language from observing the structure rules (again, this answer is not significantly more frequent than all other options in the non-generative group).

6.3.2.6 Data and theoretical virtues

There was a significant association between theoretical orientation and answers to this question ($p = .03, V = .22$). The non-generativists were more likely to agree that having reliable data is more important than whether one's theory lives up to theoretical virtues, whereas generative participants were more likely to choose the option *neither agree nor disagree* than non-generative participants (this was not the majority position in the generative group).

6.3.2.7 Traditional methods:

There was a significant association between answers to this question and theoretical orientation ($p < .001, V = .38$). The non-generative participants were more likely than generative participants to disagree with the statement that traditional methods provide good evidence, while generative participants were more likely to agree with this than non-generative participants.

6.3.3 Specialisations

In this subsection, I go through the significant associations with each specialisation listed in the questionnaire. The results are presented question by question.

As mentioned above, tests were performed both for the group of all participants, for the large generative group, for the small generative group, and for the non-generative group. For each result, I will note the relevant group.

6.3.3.1 Origin – experience or competence

See table 6.67 for an overview of the results. Within the group of all participants, there was a significant association with specialisation in syntax for this question ($p = .001, V = .32$). The syntacticians were more likely than non-syntacticians to think that linguistic intuitive judgements are good evidence because of their connection to competence.

Group	p	V	Specialisation	Answer
All	.001	.32	Syntax	Competence, Other
			Not syn	Reflections
Non	.041	.32	Syntax	Other
			Not syn	-
All	.010	.25	Historical	Other
			Not hist	Competence
All	.013	.26	Socio	Reflections
			Not socio	Competence
All	.002	.31	Cognitive	Reflections
			Not cogn	Competence, Other

Table 6.67: Associations between specialisation and Origin (Q2)

Within the non-generative group, there was a significant association with specialisation in syntax for question 2 ($p = .041, V = .32$). The syntacticians were more likely than the non-syntacticians to choose the option *other*.

Within the group of all participants there was a significant association between specialisation in historical linguistics and question 2 ($p = .010, V = .25$). The historical linguists were more likely than others to answer *other* and significantly less likely than other participants to answer that linguistic intuitive judgements are good evidence because of their connection to speakers' competence.

Within the group of all participants, there was also a significant association between specialisation in sociolinguistics and this question ($p = .013, V = .26$). The sociolinguists were more likely than non-sociolinguists to answer that linguistic intuitive judgements make good evidence because they are speakers' reflections about language.

Among all participants, there was further a significant association between specialisation in cognitive linguistics and this question ($p = .002, V = .31$). The cognitive linguists were more likely than other participants to answer that linguistic intuitive judgements make good evidence because they are speakers' reflections about language.

6.3.3.2 The role of the mental grammar

Within the large generative group, there was a significant association between specialisation in semantics and the question about the role of the mental grammar ($p = .012, V = .39$). The same was the case within the small generative group ($p = .008, V = .45$). For both groups, the semanticists were more likely than non-semanticists to think that it is a poor description of how linguistic intuitive judgements are formed to say that they are deduced from the mental grammar of speakers. See table 6.68 for an overview.

Group	p	V	Specialisation	Answer
L gen	.012	.39	Semantics	Poor description
			Not sem	-
S gen	.008	.45	Semantics	Poor description
			Not sem	-

Table 6.68: Associations between specialisation and Deduced (Q4)

6.3.3.3 Fallibility of linguistic intuitive judgements

See table 6.69 (page 165) for an overview of the results. Within the group of all participants, there was a significant association with specialisation in syntax and this question ($p = .020, V = .24$). The syntacticians were more likely than non-syntacticians to think that linguistic intuitive judgements could not be mistaken if we could abstract away performance factors.

Within the group of all participants, there was a significant association with specialisation in corpus linguistics and question 5 ($p = .019, V = .24$). The corpus linguists were significantly more likely than non-corpus linguists to agree that linguistic intuitive judgements could be mistaken even if performance factors were filtered out and significantly less likely than others to disagree with this.

Within the small generative group, there was a significant association between specialisation in theoretical linguistics and this question ($p = .041, V = .34$). The theoretical linguists were less likely than others to answer *neither agree*

nor disagree to the question about whether linguistic intuitive judgements could be mistaken if performance factors were filtered out.

Within the group of all participants, there was a significant association between specialisation in typology and question 5 ($p = .012, V = .26$). The typologists were more likely than non-typologists to answer *neither agree nor disagree*. They were significantly less likely than others to answer *disagree*.

Group	p	V	Specialisation	Answer
All	.020	.24	Syntax	Disagree
			Not syn	-
All	.019	.24	Corpus	Agree
			Not corp	Disagree
S gen	.041	.34	Theoretical	-
			Not theo	Neither
All	.012	.26	Typology	Neither
			Not typo	Disagree

Table 6.69: Associations between specialisation and Fallibility (Q5)

6.3.3.4 The subject matter of grammatical research

For an overview of results, see table 6.70. Within the group of all participants, there was a significant association with specialisation in syntax ($p = .001, V = .32$). The syntacticians were more likely than non-syntacticians to think that we should be interested in the linguistic capacity of the mind.

Group	p	V	Specialisation	Answer
All	.001	.32	Syntax	Capacity
			Not syn	Patterns
All	.004	.28	Corpus	Patterns
			Not corp	Capacity
Non	.010	.37	Psycho	Other
			Not psycho	Patterns

Table 6.70: Associations between specialisation and Subject matter (Q6)

Within the group of all participants, there was a significant association with specialisation in corpus linguistics and answers to this question ($p = .004, V = .28$). The corpus linguists were more likely than non-corpus linguists to think that we should ultimately be interested in the abstract patterns in linguistic data.

Within the group of non-generative participants, there was a significant association between specialisation in psycholinguistics and question 6 ($p = .010, V = .37$). The psycholinguists were more likely than non-psycholinguists to answer *other* and significantly less likely to answer that we should ultimately be interested in the patterns in behaviour.

6.3.3.5 Structure rules and mental rules

See an overview of results in table 6.71. There was a significant association between specialisation in morphology and answers to question 7 within the large generative group ($p = .007, V = .42$). The morphologists were more likely than the non-morphologists to answer that we cannot infer anything about how the mind processes language based on structure rules.

Within the non-generative group, there was a significant association between specialisation in semantics and this question ($p = .046, V = .37$). The semanticists were more likely than non-semanticists to answer that we cannot infer anything about how the mind processes language from the structure rules we observe.

Group	p	V	Specialisation	Answer
L gen	.007	.42	Morphology	Cannot infer anything
			Not morpho	-
Non	.046	.37	Semantics	Cannot infer anything
			Not sem	-

Table 6.71: Associations between specialisation and Rules (Q7)

6.3.3.6 The form of the implementation of mental rules

Within the small generative group there was a significant association between specialisation in theoretical linguistics and this question ($p = .022, V = .60$). Note that there were only 9 theoretical linguists in this group. The theoretical linguists were more likely than others to answer that we should not think of the rules in speakers' minds as being represented. They were also significantly less likely to choose the option *represented*. See table 6.72 for an overview.

Group	p	V	Represented	Answer
S gen	.022	.60	Theoretical	Not represented
			Not theo	Represented

Table 6.72: Associations between specialisation and Implementation (Q8)

6.3.3.7 Data and theoretical virtues

Among all participants, there was a significant association between specialisation in semantics and the question about reliable data and theoretical virtues ($p = .046, V = .21$). The semanticists were significantly less likely than non-semanticists to agree that reliable data is more important than theoretical virtues. They were not significantly more likely to choose any of the options that the non-semanticists were. See table 6.73 for an overview.

Group	p	V	Specialisation	Answer
All	.046	.21	Semantics	-
			Not sem	Agree

Table 6.73: Associations between specialisation and Theoretical virtues (Q10)

6.3.3.8 Gradience in grammar

See table 6.74 for an overview of the results. Within the group of all participants, there was a significant association with specialisation in syntax and answers to question 11 ($p = .010, V = .26$). The syntacticians were more likely than non-syntacticians to answer *other*.

Among all participants, there was a significant association between specialisation in theoretical linguistics and this question ($p = .014, V = .25$). The theoretical linguists were more likely than non-theoretical linguists to think that gradience in linguistic intuitive judgements most likely comes from extra-grammatical phenomena.

Group	p	V	Specialisation	Answer
All	.010	.26	Syntax	Other
			Not syn	Degrees
All	.014	.25	Theoretical	Extra-grammatical
			Not theo	Degrees

Table 6.74: Associations between specialisation and Gradience (Q11)

6.3.3.9 Traditional methods

For an overview of results, see table 6.75 (page 168). Among all participants, there was a significant association between specialisation in theoretical linguistics and this question ($p = .014, V = .26$). The theoretical linguists were more likely than non-theoretical linguists to agree that traditional methods provide good evidence.

Within the large generative group, there was a significant association between specialisation in pragmatics and answers to the question about traditional methods ($p = .026, V = .36$). The same was the case within the small generative group ($p = .029, V = .36$). Within both groups, pragmaticists were more likely than others to answer *neither agree nor disagree*.

Group	p	V	Specialisation	Answer
All	.014	.26	Theoretical	Agree
			Not theo	Disagree
L gen	.026	.36	Pragmatics	Neither
			Not prag	-
S gen	.029	.36	Pragmatics	Neither
			Not prag	-

Table 6.75: Associations between specialisation and Traditional methods (Q12)

6.3.3.10 Experimental methods

For an overview of results, see table 6.76. Among all participants, there was a significant association between specialisation in cognitive linguistics and answers to question 13 ($p = .045, V = .22$). The cognitive linguists were more likely than others to agree that experimental methods should be used to collect and analyse linguistic intuitive judgements.

Among all participants, there was a significant association between specialisation in theoretical linguistics and answers to this question ($p = .034, V = .22$). The theoretical linguists were significantly less likely than non-theoretical linguists to agree that experimental methods of collection and analysis should be used. They were not significantly more likely to choose any option than other participants.

Group	p	V	Specialisation	Answer
All	.045	.22	Cognitive	Agree
			Not cogn	-
All	.034	.22	Theoretical	-
			Not theo	Agree

Table 6.76: Associations between specialisation and Experimental methods (Q13)

6.3.3.11 Significance 2

Within the large generative group, there was a significant association between specialisation in pragmatics and answers to the question of whether researchers should set aside time to think about the issues discussed in the questionnaire ($p = .041, V = .33$). The same was the case within the small generative group ($p = .038, V = .39$). In both groups, pragmaticists were more likely than others to disagree that researchers should set aside time to consider why linguistic intuitive judgements can serve as evidence. Pragmaticists in the small generative group were also significantly less likely to agree to this than other other participants. See table 6.77 for an overview.

Group	p	V	Specialisation	Answer
L gen	.041	.33	Pragmatics	Disagree
			Not prag	-
S gen	.038	.39	Pragmatics	Disagree
			Not prag	Agree

Table 6.77: Associations between specialisation and Significance 2 (Q15)

6.3.3.12 Other specialisations

There were no significant associations between answers to questions and the following specialisations: phonetics ($n = 11$ among all participants; $n = 2$ in the large generative group), phonology ($n = 18$; $n = 9$), anthropological linguistics ($n = 6$; $n = 1$; one significant result was excluded because of the low number of participants in the relevant group), language acquisition ($n = 31$; $n = 18$), text linguistics ($n = 6$; $n = 3$; one significant result was excluded because of the low number of participants in the relevant group). The additional specialisations specified by participants were not frequent enough that this analysis was thought to be useful (see the overview of participants and their specialisations in section 5.5).

6.3.4 Methodology: experimental or traditional

The results presented in this subsection are based on the experimentalist-traditionalist scale described above in section 6.3.1. The analysis was carried out both on the group of all participants, the large generative group, and the non-generative group (there were no significant results within the last group). See an overview of all results presented in this section in table 6.78.

Question	Group	p	V	Experimentalists	Traditionalists
Origin (Q2)	All	.007	.27	Reflection	Competence
Rules (Q7)	All	.024	.24	Cannot infer anything	Good hypothesis
Experts (Q9)	All	.002	.27	Worse	Better
	L gen	$< .001$	.37	Worse	Better
Theoretical virtues (Q10)	All	$< .001$	.31	Agree	Neither Disagree
	L gen	.006	.34	Agree	Disagree

Table 6.78: Overview of associations with stance on methodology

6.3.4.1 Origin – experience or competence

There was a significant association within the group of all participants between stance on methodology and the answers to question 2 ($p = .007, V = .27$). The experimentalists were more likely to think that linguistic intuitive judgements make good evidence because they reflect speakers' reflections about language, whereas the traditionalists were more likely to think that it is because of the connection between linguistic intuitive judgements and competence.

6.3.4.2 Structure rules and mental rules

Within the group of all participants, there was a significant association between answers to this question and stance on methodology ($p = .024, V = .24$). The experimentalists were more likely to answer that we cannot infer anything about how language is processed by looking at structure rules. Traditionalists were more likely to say that it is a good hypothesis that structure rules are implemented in the minds of speakers.

6.3.4.3 Experts and ordinary speakers

There were significant associations between answers to this question and participants' stance on methodology, both within the group of all participants ($p = .002, V = .27$) and within the large group of generative participants ($p < .001, V = .37$). Within both groups, the experimentalists were more likely to answer that experts make worse subjects than lay subjects, whereas the traditionalists were more likely to answer that experts are better subjects than lay people.

6.3.4.4 Data and theoretical virtues

There were significant associations between answers to this question and stance on methodology both within the group of all participants ($p < .001, V = .31$) and within the large group of generative linguists ($p = .006, V = .34$). In both groups, the experimentalists were more likely than traditionalists to agree that having reliable data is more important than theoretical virtues. Traditionalists in the group of all participants were more likely than experimentalists in that group to answer *neither agree nor disagree*, and in both groups the traditionalists were more likely to disagree that reliable data is more important than theoretical virtues (although these were minority positions in both the large generative group and the group of all participants).

6.3.5 Summing up

In the following, I sum up the main results of the association analyses. In the next chapter (chapter 7), I comment on and discuss these results as well as the results presented in the previous parts of this chapter. Note again that the mentioned positions are *not* necessarily the majority positions within each group, rather they indicate areas where participants from one group were more likely to give a particular answer than participants from the other group.

Within the generative group, there was not much of a difference in how questions were answered between those who identified solely as generative and those who mentioned generative/formal linguistics as one of their influences. Comparing the large generative group to the non-generative group, however, there were a number of differences. The generative group was more likely than the non-generative group to take the following positions: that we should study the mental linguistic capacities rather than the abstract patterns in the data, that competence is what makes linguistic intuitive judgements good evidence, that it is a good hypothesis that the structure rules of grammar are implemented in the minds of speakers, that linguistic intuitive judgements are actually deduced from speakers' mental grammars, that linguistic intuitive judgements could not be mistaken if performance factors were filtered out, that traditional methods provide good evidence for grammars, and that it is not the case that reliable data is more important than theoretical virtues.

The non-generative group, on their side, was more likely than the generative group to take the following positions: that we should study the abstract patterns found in linguistic data, that linguistic intuitive judgements are good evidence because they are speakers' reflections about language, that we cannot infer anything about how the mind processes language based on structure rules, that it is a poor description to say that linguistic intuitive judgements are deduced

from the mental grammar, that linguistic intuitive judgements could be mistaken even if performance factors were filtered out, that traditional methods do not provide good evidence for grammatical theories, and that reliable data is more important than theoretical virtues. These differences between the groups are hardly surprising, but as I will discuss in chapter 7, it may be interesting to look at the differences between what the majority view is within a group and what answers were associated with membership of that group.

There were a number of significant associations with areas of specialisation as well. Here I will just mention a few interesting results. On the areas where the syntacticians' answers differed significantly from those of non-syntacticians, the syntacticians were more likely to answer like the generative group as described above. On the other hand, semanticists, corpus linguists, sociolinguists, and cognitive linguists differed from their respective counterparts in similar ways to how the non-generativists differ from the generativists (note that most of these groups only differ from their counterparts on one or two issues though). These trends are found within the group of all participants and sometimes within other groups as well. So, interestingly, we not only find differences between theoretical orientations but also somewhat similar divisions among specialisations within linguistics with syntacticians being more like the generative group and semanticists, corpus linguists, sociolinguists, and cognitive linguists being more like the non-generative group.[23] These results will be discussed again in chapter 7.

Another interesting result to look into concerns the theoretical linguists. Specialisation in theoretical linguistics was, not surprisingly, significantly associated with answering that traditional methods provide good evidence for theories of grammar and that gradience in intuitive judgements is most likely due to extra-grammatical factors (both of these results were found within the group of all participants). More interestingly, within the small generative group among those who answered that structure rules are likely to be implemented in the minds of speakers, specialisation in theoretical linguistics was also associated with answering that these rules are most likely not represented (note that only 20 participants answered this question as it was contingent on the answer given to the previous question; of these 20, 9 were theoretical linguists and 11 were not). So, the generative theoretical linguists were less likely than non-theoretical generative linguists to think that, if we take structure rules to be implemented in the minds of speakers, those rules are likely to be *represented* in the minds of speakers. This, at least to me, is somewhat surprising.

The significant associations with stance on the experimentalist-traditionalist

[23]This is perhaps not surprising, taking into consideration that generative linguistics has long given syntax a very prominent place in the research program, whereas non-generative approaches to linguistics tend to give more weight to social and non-language specific cognitive explanations than generative linguistics traditionally does (see chapter 4)

debate were perhaps not surprising. On the issues that are relevant to the etiology debate, the traditionalists were more likely than experimentalists to take the following positions: that intuitive judgements are good evidence because of their connection to competence and that it is a good hypothesis that structure rules are implemented in the mind. On the other hand, experimentalists were more likely than traditionalists to think the following: that linguistic intuitive judgements are good evidence because they are speakers' reflections about language and that we cannot infer anything about how the mind processes language from structure rules. When looking just at the associations with answers to questions from the Experimental Syntax debate, we see that on two out of three issues, the experimentalists and traditionalists are more likely to answer what we would expect: Experimentalists are more likely to answer that experts make worse subjects than lay people and that reliable data is more important than theoretical virtues, and traditionalists are more likely to answer that experts make better subjects and to disagree that reliable data is more important than theoretical virtues. On the issue of gradient grammar, there were no significant associations with stance on methodology.

Chapter 7

Summary and discussion

As laid out in section 5.1, the main questions that motivated the study described in the two previous chapters were the following:

Is it the case that a majority of linguists working in the generative framework subscribe to the Voice of Competence view (VoC) as characterised by Michael Devitt? On that view, linguistic intuitive judgements are good evidence for grammatical theories because the content of these judgements is derived fairly directly from the rules of the speaker's grammar, which are implemented in the speaker's mind. Or is it rather the case that a majority of generative linguists hold one of the other views put forward in the etiology debate, such as the Critics' Views? On that view, it is the connection to speakers' competence that makes linguistic intuitive judgements good evidence, but it does not require that intuitive judgements get their informational content directly from the speaker's mental grammar. Or do they rather hold the Modest Explanation, on which intuitive judgements are central-processor judgements with no special input from the speaker's linguistic competence? Instead, on that view, intuitive judgements are based on the speaker's experience with their language and the (folk) theoretical concepts they apply when making the judgements (see chapter 2 for more details on the three views). And what about non-generative linguists? Is there a difference in how generative and non-generative linguists view this question?

Another main question behind this study was what the majority positions within generative and non-generative linguistics respectively are about the main issues of the Experimental Syntax Debate. Should we prefer to use the intuitive judgements of experts or rather those of lay subjects as evidence for theories of grammar? Should we use experimental methods to collect and analyse linguistic intuitive judgements, or does the informal armchair provide good evidence? Is it more important to ensure that one's data is reliable than to make sure that one's theory lives up to theoretical virtues such as simplicity and elegance? And

how should we interpret the gradience that we see in some intuitive judgement experiments: Is it likely to be due to the (graded) nature of speakers' grammars, or is it more likely to be an effect of extra-grammatical factors? (see chapter 3 for more details). And do generative and non-generative linguists differ on these issues?

An additional question was whether there is a difference between the experimentalists and traditionalists of the Experimental Syntax debate when it comes to their views regarding the etiology question. And, finally, is there an association between particular linguistic specialisations and particular views on the etiology question and the Experimental Syntax debate?

In this chapter, I summarise and discuss the results from the analyses presented in the previous chapter to answer the questions above.

7.1 The etiology debate

First, let us look at the results that are relevant to the etiology debate. I start out with the answers from the generative group, then look at the answers from the non-generative group, and then I move on to the results from the association analyses.

7.1.1 The majority generative view

Within the group of generative participants, there was no majority for the views that make up VoC. In other words, it does not seem to be the case that VoC is the majority view in generative linguistics as debated following Devitt's (2006c) attribution of the view to generative linguists. However, there was also no majority for the alternative competence-based views proposed by Devitt's critics. Finally, there was no majority for the Modest Explanation among generative linguists either.

Instead, the view that did emerge as the majority view among generative linguists in this survey can be characterised as follows: It is a competence-based view, i.e., we can rely on intuitive judgements as evidence for theories of grammar because they reflect speakers' linguistic competence. Furthermore, they do so relatively directly: If we could completely filter out the influence of performance factors, speakers' intuitive judgements would always correctly reflect the grammatical status of sentences (though recall that we should only tentatively accept this result, see section 6.1.2.4). So far, the majority generative view looks a lot like VoC.

One might think that such a relatively direct route from the speaker's competence to their intuitive judgements could only be explained by the content of

intuitive judgements being, somehow, directly derived from the speaker's mental grammar (with performance factors filtered out). But the majority of generative participants rejected such a direct path between the speaker's mental grammar and the content of intuitive judgements. A majority also rejected the idea that what we usually think of as the rules of a speaker's language (its structure rules) are even themselves necessarily implemented in speakers' minds, which seems like a prerequisite for the content of intuitive judgements being deduced from those rules. But if the rules of a speaker's language are not themselves implemented in speakers' minds, and if the content of intuitive judgements is not derived from the speaker's mental grammar, then how can one explain that intuitive judgements would always correctly reflect the true grammaticality status of a sentence if only performance factors could be filtered out?

The present study unfortunately cannot tell us more about how generative participants reconcile this, but it would be a highly interesting question to pursue in future, perhaps more qualitatively focused studies. One potential solution is that the speaker's central processor is involved, which would mean that intuitive judgement are not "deduced" directly from the speaker's mental grammar, but that the involvement of the central processor is so relatively minor that, in practice, the resulting intuitive judgements can be said to be infallible, performance factors aside. This, I believe, is the position of Rey (2013, forthcoming), and it will be debated further in chapter 8. It is, of course, also possible that the majority view on the etiology question among generative linguists is incoherent. Further studies could probe these possibilities.

On two further issues, the majority generative view distinguishes itself from VoC and from the Critics' Views and the Modest Explanation as well. On VoC, as well as on the Critics' Views, the kind of intuitive judgements that linguists should use as evidence for their grammatical theories are acceptability judgements. On the Modest Explanation, it is grammaticality judgements. On the majority generative view, there is no such clear agreement. In fact, there are roughly equally large groups arguing for both positions, and there is even a sizeable group that maintains that *both* judgements of acceptability and grammaticality can be used as evidence for grammatical theories.

The same was the case for the issue of what, ultimately, we are interested in when we study grammatical phenomena: the abstract patterns found of linguistic behaviour or the underlying mental, linguistic capacities. Both VoC and the Critics' Views take the mentalist stance that the study of grammar is ultimately about speakers' mental linguistic capacities. On the Modest Explanation, on the other hand, the aim of grammatical research is to describe the non-mental linguistic reality. The majority generative view found in this study again differs from both these views as there were large groups of participants arguing for

either side, as well as a third group rejecting that one of these aims ultimately has priority over the other.

These last two results were surprising, as the generative literature in general seems to advocate that only acceptability judgements can be used as evidence and that the study of our mental linguistic capacities should have priority over the study of the patterns found in linguistic behaviour.

All the all, the majority generative view on the etiology questions is a competence-based view, however, it is one that differs from both VoC and the Critics' Views.

7.1.2 The majority non-generative view

Among the non-generative participants in the study, there was also no clear-cut support for VoC or the Critics' Views. However, the majority non-generative view that emerged from the analyses is fairly close to the Modest Explanation. The main difference between the majority non-generative view and the Modest Explanation is that on the former, we should use intuitive judgements of acceptability rather than grammaticality as evidence for theories of grammar.

On the majority non-generative view, linguistic intuitive judgements make good evidence for grammatical theories because they express speakers' reflections about language, and it is not the case, on this view, that if we could filter out performance factors from intuitive judgements that they would have to be correct about the grammaticality status of sentences. On this view, the ultimate goal of grammatical research is to study the externally observable patterns in linguistic behaviour. So far, this view is consistent with the Modest Explanation.

On two issues, there was no clear majority among the non-generative participants. However, from the way the answers patterned the following was clear: It is not part of the majority non-generative view that structure rules are implemented in the minds of speakers, and, similarly, it is not part of this view that intuitive judgements are deduced from speakers' mental grammars.

This view is clearly not competence-based in the same way as the generative majority view and the Critics' View, and it is very dissimilar from VoC. It is, however, so far in line with the Modest Explanation. But, as mentioned, on the majority non-generative view, only intuitive judgements of acceptability (rather than intuitive judgements of grammaticality) can serve as evidence for grammatical theories. This marks a difference with the Modest Explanation.

How does this difference between relying on grammaticality judgements and relying on acceptability judgements affect the view? For instance, does it add up to a coherent view to say that intuitive judgements are based on speakers'

experience with and reflections on language (rather than competence) and to say that the intuitive judgements we should use as evidence are judgements of acceptability? This study does not answer how non-generative linguists would reconcile those ideas, but one potential solution could be the following. Speakers will have experiences of sentences being acceptable or not to them from their daily linguistic lives. If they become aware of this and reflect on it, they might be able to use that category of experience to judge new sentences against when asked for a linguistic intuitive judgement. That way, acceptability rather than grammaticality would be the relevant type of judgement, but the judgement itself would still be due to the speaker's experience with and reflection on language rather than their competence directly. For that reason, judgements would be fallible even if performance factors were filtered out, as speakers' reflections on their experience with acceptability might be flawed for reasons not directly to do with performance factors. Whether this is indeed a part of the non-generative majority view would be an interesting question to take up in further investigations. For now, I will just mention that if the proposed explanation is in fact part of the majority non-generative view, this view would be fairly close to the account that I propose and defend in chapter 8.

7.1.3 Associations

7.1.3.1 Generative and non-generative linguists

We have already seen in the two preceding sections that the majority view among the generative participants of the study differs from the majority view of the non-generative participants. We saw from the association analysis in section 6.3 as well that for all questions in the etiology debate, save the questions Acceptability/grammaticality (Q3) and Implementation (Q8), participants' theoretical orientation (generative or non-generative) was significantly associated with different answers to etiology questions. Table 7.1 (page 179) shows what answers the two theoretical orientations were associated with respectively.[1]

From table 7.1, we see that in some cases, one theoretical orientation is associated with an answer that was not the majority view within that group. That just means that participants in this theoretical group was significantly more likely to choose this answer than participants in the other theoretical group. For instance, in the question about whether the ultimate goal of doing grammatical research should be to understand the abstract patterns found in linguistic data or rather to understand the underlying mental linguistic capacities, generative participants were more likely than non-generative linguists to say that we should

[1]This table is a slightly modified version of table 6.66 from section 6.3.2 (the results that are relevant to the Experimental Syntax debate have been left out).

Question	p	V	Generative	Non-generative
Origin (Q2)	$< .001$	.43	Competence	Reflections
Deduced (Q4)	.001	.34	Actual process	Poor description
Fallibility (Q5)	$< .001$	.42	Disagree	Agree, Neither
Subject matter (Q6)	$< .001$	.35	Capacity	Patterns
Rules (Q7)	$< .001$	.36	Good hypothesis	Cannot infer anything

Table 7.1: Associations between etiology views and theoretical orientation, all participants

be interested in the underlying mental capacities even though there was not a significant majority behind this answer within the generative group.

Some of the positions which seem prominent in the generative literature were not found to be majority generative positions, for instance, the view that structure rules are themselves implemented in the minds of speakers, which I agree with Devitt seems from the literature to be (or at least at one time to have been) a prominent view within generative linguistics. Perhaps the results of the association analysis can shed some light on why these views seem to stand out in the generative literature even though they appear not to be the majority generative view. It could be that these views owe some of their prominence in the literature to the fact that they are the views that differentiate the main two camps of linguistic approaches, generative and non-generative linguistics. In the question of whether intuitive judgements are deduced from the speaker's mental grammar, the generative linguists were more likely than the non-generative linguists to answer that this is the actual process of how intuitive judgements are formed, and the non-generative linguists, on their side, were more likely than the generative linguists to answer that this is a poor description of that process (though neither of those views were the majority view within each group). In the question of whether we should think that structure rules are implemented in the minds of speakers, generative linguists were more likely than their non-generative colleagues to answer that this is a good hypothesis, whereas the non-generative linguists were more likely than the generative linguists to think that we cannot infer anything about how the mind processes language from observing the structure rules of languages (again, neither of these answers were the majority position in the respective groups).

In other words, VoC might not be the majority view among generative linguists, but it seems that generative linguists are more likely than non-generative linguists to hold most of the central component views that go into VoC. This could be part of the explanation why VoC (or some view like it) seems to crop up in the generative literature, even if it might not be the current majority view.

7.1.3.2 Experimentalists and traditionalists

Were there any differences between how experimentalists and traditionalists (see section 6.3.4) viewed the questions of the etiology debate? Not for all questions, but there were two significant associations between views in the etiology debate and stance on methodology. Table 7.2 shows these result.[2]

Question	p	V	Experimentalist	Traditionalist
Origin (Q2)	.007	.27	Reflection	Competence
Rules (Q7)	.024	.24	Cannot infer anything	Good hypothesis

Table 7.2: Associations between etiology views and stance on methodology, all participants

The experimentalist participants were more likely than the traditionalists to think that intuitive judgements can serve as evidence for grammatical theories because they are speakers' reflections about language. They were also more likely than traditionalists to think that we cannot infer anything about how the mind processes language from observing structure rules. The traditionalists, on the other hand, were more likely than experimentalists to think that intuitive judgements cam serve as evidence because they express speakers' competence and that it is a good hypothesis that structure rules are implemented in the minds of speakers.

On these two issues, the experimentalists were more likely to choose answers in line with the Modest Explanation, while traditionalists were more likely to choose answers in line with VoC. These results show that there are some differences in etiology views between those who prefer experimental methods and those who prefer traditional methods, but even so, on most issues in the etiology debate there were no significant differences. Overall, it does not look like the etiology debate and the Experimental Syntax debate divide participants in this study according to the same lines.

7.1.3.3 Different linguistic specialisations

There were a number of differences in views on the etiology debate associated with different specialisations within linguistics.[3]

Looking first at the results found within the group of all participants (both the generative group and the non-generative group), syntacticians were more

[2]This table is adapted from table 6.78 in section 6.3.4. The adapted table shown here only presents the results relevant to the etiology debate and leaves out those results relevant to the Experimental Syntax debate.

[3]In the following, I leave out the cases where linguists of a particular specialisation were more likely than other participants to answer *other* or *neither*.

likely than non-syntacticians to think that intuitive judgements can serve as evidence for grammatical theories because they express speakers' competence, that intuitive judgements would always be correct if only performance factors could be filtered out, and that we should ultimately be interested in the mental linguistic capacity rather than the abstract patterns in linguistic behaviour.

Sociolinguists and cognitive linguists were more likely than other participants to think that intuitive judgements can serve as evidence because they are speakers' reflections about language. Corpus linguists were more likely than other participants to think that linguistic intuitive judgements could be incorrect about the grammatical status of sentences even if we could filter out performance factors, as well as to think that we should ultimately be interested in the abstract patterns found in linguistic behaviour rather than in the underlying mental capacity.

As mentioned in section 6.3, these results share a certain similarity with the results found for the generative and non-generative groups respectively, with syntacticians answering more like generative linguists, and sociolinguists, cognitive linguists, and corpus linguists answering more like non-generative linguists. This could, as mentioned in section 6.3, be due to the fact that generative and non-generative linguistics respectively traditionally has given different levels of prominence to syntax vis-a-vis sociolinguistics, cognitive linguistics, and corpus linguistics. However, it might also be due to the overall composition of the group of syntacticians, sociolinguists, cognitive linguists, and corpus linguists respectively with regards to theoretical orientation. Table 7.3 shows the number of participants with each of these specialisations that fall within the generative and non-generative groups respectively.[4]

Specialisation	Generative	Non-generative
Syntax	$n = 62$	$n = 16$
Corpus	$n = 9$	$n = 26$
Socio	$n = 1$	$n = 16$
Cognitive	$n = 1$	$n = 25$

Table 7.3: Number of generative and non-generative linguists in selected specialisations

As we can see from table 7.3, there is a prevalence of generative participants within the group of syntacticians, and similarly there is a prevalence of non-generative linguists within the groups of sociolinguists, corpus linguists, and cognitive linguists. In the case of sociolinguists and cognitive linguists, there is

[4]These numbers were also presented in table 5.4 in section 5.5.

only 1 generative linguist in each group. This distribution of participants with each respective theoretical orientation looks like a plausible answer to why syntacticians seemed to answer like generative linguists and sociolinguists, corpus linguists, and cognitive linguists seemed to answer like non-generative linguists. It does not immediately seem to be an effect of specialisation in itself, and I will leave these results out of further discussions.

In the various subgroups of participants (non-generative linguists, the small generative group, and the large generative group), the same distribution issues do not arise. Within the non-generative group, semanticists were more likely than others to think that we cannot infer anything about how the mind processes language from looking at structure rules. Within the large generative group, the morphologists were also more likely than others to answer that we cannot infer anything about how the mind processes language from looking at structure rules, and the semanticists were more likely than others to answer that "deduction talk" is a poor description of how intuitive judgements are formed. Within the small generative group, theoretical linguists who thought that structure rules are implemented in the minds of speakers were more likely than others to answer that those structure rules are not represented.

As we can see, there were only a few associations between etiology views and specialisation within the theoretical subgroups in the study. Semanticists and morphologists, both within the non-generative group and the large generative group, seem to be more likely to answer in the same way as non-generative linguists. We only see this for a number of questions though. The more interesting result, I think, is that theoretical linguists in the small generative group who thought that structure rules are implemented in speakers' minds were more likely than others to think that those structure rules are *not* represented.

This is interesting, as one might expect the theoretical linguists to be most in tune with the latest theoretical developments within the framework. Allowing for some speculation, this pattern of answers would fit if it used to be the received generative view that rules are represented in speakers' minds *and* if this position has since been abandoned. If that is the case, this new position could have taken longer to filter down to non-theoretical linguists. All this is very speculative, however, and it is another result that it would be interesting to follow up on in future studies.

All in all, it does not look as if stance on the etiology question is in general associated with specialisation within linguistics.

7.2 The Experimental Syntax debate

Let us move from the results that are related to the etiology debate to the ones that are relevant to the Experimental Syntax debate. I start with the majority generative view, then I look at the majority non-generative view, and then I look at the results from the association analyses.

7.2.1 The majority generative view

Among the generative participants, there was no clear-cut support for one side of the Experimental Syntax debate over the other. The majority of generative participants thought that reliable data should be favoured over theoretical virtues when evaluating theories and that both traditional, informal methods and experimental methods for collecting linguistic intuitive judgements are useful. There was no clear majority on the issues of whether the linguistic intuitive judgements of experts provide better evidence for theories of grammar than those of ordinary speakers, or on the issue of whether the gradience found in some intuitive judgement experiments is best explained by a grammar that is graded in nature or by extra-grammatical factors.

Of all these issues, it is only on the question of reliable data vs. theoretical virtues that we see clear support for one option over another. But as mentioned in section 6.1.3.2, the view that reliable data is more important than theoretical virtues is not exclusive to the experimentalist side of the debate (although the view that theoretical virtues are more important than reliable data does seem to be exclusive to the traditionalist side). In other words, we cannot conclude that this result counts in favour of the experimentalist side of the debate. All in all, there is no clear support for either side of the debate among the generative participants.

Perhaps this finding is not surprising. This result, with opinions divided between the experimentalist side and the traditionalist side of the debate, seems to mirror the ongoing debate between experimentalists and traditionalists within generative linguistics.

7.2.2 The majority non-generative view

The non-generative participants' views on the issues of the Experimental Syntax debate were closer to those of the experimentalists than those of the traditionalists. A majority of the non-generative linguists thought that we should prioritise reliable data over theoretical virtues, that the gradience found in some intuitive judgement experiments is really due to gradience in the grammar, and that, in general, experimental methods should be used to collect and analyse linguistic

intuitive judgements. There was no clear majority on the issues of whether the linguistic intuitive judgements of experts provide better evidence for theories of grammar than those of ordinary speakers, or about whether informally collected intuitive judgements make good evidence or not.

Still, overall, the majority view among the non-generative participants was fairly close to the experimentalist views put forward in the Experimental Syntax debate.

7.2.3 Associations

7.2.3.1 Generative and non-generative linguists

As the two sections above show, we did see different views on the Experimental Syntax issues between participants working within generative linguistics and participants working within non-generative linguistics. Table 7.4 shows the issues on which there were significant associations between stance on the methodology questions and participants' theoretical orientation.[5]

Question	p	V	Generative	Non-generative
Theoretical virtues (Q10)	.032	.22	Neither	Agree
Traditional methods (Q12)	$< .001$	.38	Agree	Disagree

Table 7.4: Associations between Experimental Syntax views and theoretical orientation, all participants

In table 7.4, we see that there were only significant associations between participants' stances on the Experimental Syntax questions and their theoretical orientations in two cases. Non-generative linguists were more likely than generative linguists to think that having reliable data is more important than whether one's theory lives up to theoretical virtues, and they were also more likely than generative linguists to not think that the traditional armchair method provides good evidence for theories of grammar. The generative linguists, on their side, were more likely to think that the traditional method does provide good evidence.

These results partly mirror the ones laid out in the sections above, namely that while there seems to be a consensus within the non-generative group in favour of the experimental side of the Experimental Syntax debate, opinions are more divided among the two sides of the debate within the generative group.

[5]This is a version of table 6.66 from section 6.3.2. The results that are relevant to the etiology debate have been left out.

Again, this is perhaps not surprising, given that there is an ongoing debate between experimentalists and traditionalists within generative linguistics.

7.2.3.2 Different linguistic specialisations

There were a few significant associations between participants' views on the Experimental Syntax issues and their specialisations in linguistics.[6]

In the group of all participants, the theoretical linguists were, perhaps unsurprisingly, more likely than other participants to think that gradience in judgement experiments is due to extra-grammatical factors and to think that traditional methods of collecting intuitive judgements yield good data. Also in the group of all participants, cognitive linguists were more likely than other participants to think that one should use experimental methods when collecting and analysing intuitive judgements.

As discussed above, the number of non-generative cognitive linguists ($n = 25$) compared to the number of generative cognitive linguists ($n = 1$) means that for the latter result, we cannot disentangle the effect of the participants' specialisation from their theoretical orientation. But what about the theoretical linguists? 35 participants reported theoretical linguistics as one of their main specialisations. Of these, 29 were in the generative group and 5 were in the non-generative group. Again, we cannot conclude that the significant associations between specialisation in theoretical linguistics and traditionalist answers to the Experimental Syntax questions are due to specialisation alone, as the results might be influenced by participants' theoretical orientation as well.

7.3 Concluding remarks

This study gives a snapshot of the current views of generative as well as non-generative linguists on two debates that centre on the use of linguistic intuitive judgements as evidence for theories of grammar: the etiology debate and the Experimental Syntax debate.

Starting with the Experimental Syntax debate, we saw, perhaps unsurprisingly, that the non-generative linguists in the study overall agreed more with the experimentalist side of the debate than with the traditionalist side. This echoes some of the worries about the use of intuitive judgements as evidence for grammatical theories mentioned in chapter 4. Among the generative linguists in this study, on the other hand, there was no clear majority support for either side of the debate. This, perhaps, reflects the state of the debate within generative

[6] In the following, I leave out the cases where linguists of one specialisation were more likely than other participants to answer *other* or *neither*.

linguistics at the moment, where a group of experimentalists are pushing for the use of more experimental methods while a group of traditionalists are arguing that the armchair method is working just fine.

As for the etiology debate, the study answers what has been an open question in the debate between Devitt and his critics: whether VoC is the majority view among generative linguists or not. The answer, as we saw above, is that it is not. However, neither of the other two main views defended in the debate are held by a majority of generative linguists either. Instead, the majority view among the generative participants in this study is a new competence-based view of its own. We also saw that the majority view among the non-generative linguists came very close to the Modest Explanation.

As mentioned, this answers one of the main questions of the debate between Devitt and his critics. However, it, of course, does not answer the other main issue of that debate, namely the normative question of which view one should prefer. I give my take on that issue in chapter 8.

Part III

Justification

Chapter 8

Justifying using linguistic intuitive judgements as evidence

8.1 Introduction

The previous chapters dealt with, among other things, the question of what the received view on the justification question is in both generative and non-generative circles in linguistics. In this chapter, I turn to the justification question itself. What makes these intuitive judgements good evidence for grammars in linguistics – why we are justified in using linguistic intuitive judgements as evidence?

In this chapter, I will first give a short summary of the current debate on the justification question, which was presented in chapter 2. I then introduce and defend a novel proposal that accounts for the evidential role of linguistic intuitive judgements and accommodates important insights from both sides of the debate. This account gives a central role to the speaker's competence in a way that is consistent with linguistic intuitive judgements being theory-laden, central processor judgements, like on Devitt's Modest Explanation (presented in chapter 2; see also section 2.1 for an introduction to the term "central processor"). I will suggest that Barbara J. Luka's (2005) proposal of affective evaluations accompanying sentence comprehension is a plausible candidate for a link between competence and judgements. After that, I sketch the role of the central processor in forming the content of the judgement, and I then consider some objections to the proposed account. Finally, I argue that this account does not provide support for the so called expertise defence, i.e., the claim that we should

use the intuitive judgements of experts rather than those of lay subjects, a claim which the Modest Explanation does support.

8.2 The current debate

Since the publication of *Ignorance of Language*, Devitt (2006b,c), Rey (2013), Maynes and Gross (2013), Ludlow (2011), and Collins (2008), among others, have been in an at times intense discussion on the topic of the etiology of linguistic intuitive judgements. The debate has two main themes: 1) Can the etiology of intuitive judgements justify their evidential use in linguistics, and, if so, how? (as mentioned in chapter 2, this is one take on the justification question.) 2) What is the received view on the etiology question among generative linguists? The second issue was the centre of parts of chapter 6. In this chapter, I focus on the main issues of the debate on the justification question itself.

In the following, I will briefly re-introduce the three central views from the etiology debate (as presented in chapter 2). In doing so, I will highlight five issues that are important to one or more of these views and which are central to the ways the views disagree with each other:

1. whether the speaker's competence provides the informational content of the intuitive judgement (or, alternatively, provides data for the central processor to process further),

2. whether linguistic intuitive judgements receive any *special* input from the speaker's competence or not,

3. the distinction between a mental state *having* a certain property and it *representing* a certain property,

4. whether expertise in linguistics improves one's intuitive judgements, and, finally,

5. the distinction between hearing a sentence *as* part of one's language and *judging* that it is part of one's language.

The debate took off after Devitt (2006c) characterised what he found to be the received view on the justification question in generative linguistics. He called this view the Voice of Competence view (VoC).[1] A crucial aspect of this view is that the speaker's linguistic competence provides the *informational content* of the judgement without influence from the central processor (issue 1). Based

[1] Although Rey uses they term "the Voice of Competence" to refer to his own view which, as we shall see, differs from what Devitt calls VoC, in the following, I use "VoC" specifically to refer to the view characterised by Devitt.

on quotes from the generative literature, Devitt describes how this could be imagined to work (calling this "the standard version"): The structure rules that linguists attempt to capture when they write grammars are represented in the minds of speakers. When a speaker forms an intuitive judgement, they derive the verdict of their judgement more or less directly from the rules in their mind with "one sort of representation leading to another" (with some room for noise, Devitt 2006c, 484). In this way, the speaker's competence does deliver special input for the judgement (issue 2). Devitt criticises this view for being immodest in the cognitive mechanisms that it relies on.[2] He also sketches another potential version (the "non-standard version") of this view where the rules are embodied in speakers' minds rather than represented, but the informational content of the judgement is still somehow fairly directly derived from these embodied rules. On my reading of Devitt's characterisation, even on the non-standard version of VoC, there is no involvement of the central processor. Devitt rejects this version of VoC as well, as he thinks that no plausible story is likely to be forthcoming about how embodied rules could issue in the content of an intuitive judgement without involvement of the central processor.

His critics think that VoC does not exhaust the possibilities of a competence-based view (see, e.g., Gross forthcoming). In this chapter, I will focus on Rey's (2013; forthcoming) account, one of the prominent and more fully developed accounts in the debate. His account is based on a view of language comprehension on which the speaker's competence produces a structural description of the comprehended sentence. For the sentence "sleepy cats like newspapers", a structural description would be a representation of (some) of the structural properties of the sentence, something like: $[_{S}[_{NP}[_{A}\text{sleepy}]\ [_{N}\text{cats}]]\ [_{VP}[_{V}\text{like}]\ [_{NP}[_{N}\text{newspapers}]]]]$. On Rey's view, these structural descriptions, which are non-conceptual and not available to the speaker at the conscious level, will be processed by the central processor and translated fairly directly into the content of the intuitive judgement. Structural descriptions of well-formed sentences will lead to the verdict "acceptable", and structural descriptions of ill-formed sentences will lead to the verdict "unacceptable".[3] Judgements about, for instance, co-reference can also be accommodated on this account since the structural description of a sentence would contain information about which items co-refer. On this view, the competence does not deliver the full informational content of the intuitive judgement (issue 1), but the competence does deliver special input to the central processor (issue 2) in the form of structural descriptions.

[2]Recall from section 2.3.6 that Devitt argues that all we need to assume is that the processing rules yield output that conforms to the structure rules.

[3]Rey is interested in judgements of acceptability rather than grammaticality, but recall that this issue is debated (see section 2.3.2). In this chapter, at least until section 8.4.2, I use the terms interchangeably.

Similar views that grant a central role to the speaker's competence are proposed by Textor (2009) and Maynes and Gross (2013), the latter developed in Gross (forthcoming) (see chapter 2).

Rey (forthcoming) emphasises the difference between the output of the competence *having* a certain property and the output of the competence *representing* that property, such as being grammatical, involving co-reference, etc. (issue 3). He argues that merely *having* a property will not entail the output being *treated as having* that property as any such mental signal might have several properties, only some of which are potentially relevant to the task at hand. For instance, he mentions the relational property of *being n minutes older than y*. A mental state might have this relational property without it being relevant for anything else to treat it as having that property. Instead, he argues, if the output of the competence *represents* the relevant properties, then it follows naturally how the central processor can *treat* the signal correctly, whereas this is not the case if the signal just *has* those properties. In case a signal just has a certain property without representing it, on his view, we need some other reason to think that the signal will be treated as having the relevant property. It is central to his account that the output from the speaker's competence, the structural descriptions, represent the relevant properties of the sentence.

On Devitt's own view, the Modest Explanation, it is the central processor, rather than the competence, that plays the central role. On this view, the subject hears a sentence and simulates the experience of attempting to comprehend or produce the sentence. Then, using themselves as a reliable guide to what a competent speaker would do, they use that experience as the datum to base a decision on. At least, this is what happens in difficult cases, whereas simpler cases might be decided without any direct reflection on a simulation (Devitt, 2010b). Contrary to the competence-centred accounts, there is no special input from the competence (issue 2). If the subject finds the sentence okay, they will likely judge it to be grammatical, and if not they might judge it to be ungrammatical or maybe grammatical but infelicitous. All this usually happens immediately and unreflectively (Devitt, 2010b). The judgement will be theory-laden, not in the sense that it is based on theoretical reasoning, but rather in the sense that all judgements are theory-laden. This ranges from the theory-ladenness of observation judgements of the type "the grass is brown" to judgements laden with folk or expert theories (Devitt, 2006c). One consequence of this is that, on the Modest Explanation, we should prefer the linguistic intuitive judgements of linguistic experts over those of lay subjects, as the experts' intuitive judgements are laden with better theories than those of lay subjects (issue 4). Devitt's view was dubbed the Modest Explanation by Textor (2009), reflecting that it only relies on cognitive processes we are already committed to.

Devitt (2013) emphasises the distinction between hearing a sentence as part of your language (which we cannot help but do as native speakers of a language) on the one hand and making the meta-linguistic judgement that a sentence is, in fact, part of your language on the other hand (issue 5). He stresses that being able to do the prior does not automatically mean that we are able to do the latter. The fact that we can be shown to be sensitive to whether a sentence is part of our language or not *in our linguistic behaviour*, Devitt thinks, does not count as evidence that we will be able to make the *meta-linguistic judgement* that the sentence is part of our language. Taking it to be such evidence would be to assume part of what needs to be argued. It is a central part of his critique against his critics that they do not pay sufficient attention to this gap. On his account, the gap is bridged by the central processor reflecting on the experience of comprehending or producing the target sentence.

As mentioned in chapter 2, each of the three proposals has been met with some criticism. VoC, while providing a clear and bold answer to the justification question, has, as Devitt argues, the problem of being cognitively immodest as it relies on cognitive assumptions that lack independent plausibility. Although some use the phrase "the Voice of Competence" when arguing for somewhat different competence-based views, to my knowledge, so far no one has come forward to defend this *particular* version of a competence-based view.

While Rey's view is similar to VoC in the sense that competence provides special input to intuitive judgements, it also differs from VoC since, on Rey's view, the competence does not provide the *informational content* of intuitive judgements, and so his view escapes the critique that it is, essentially, another version of VoC. Devitt (2013), however, argues, in effect, that if the competence does not supply the content of the judgement, then views like Rey's are basically just different versions of the Modest Explanation since "the intuitions arise from the subject's central processor reflection on the data of trying to understand the string". Recall that on VoC as Devitt characterises it, even on the non-standard version where rules are embodied rather than represented in the mind, the central processor does not have a role to play. From this reply, it seems as if Devitt intends the Modest Explanation to cover all accounts on which the central processor plays a role. But as we have seen, Rey emphasises another difference between his view and the Modest Explanation. On Rey's view, the competence provides special input for the central processor, and this is not the case on the Modest Explanation.

On the Modest Explanation, we should prefer the intuitive judgements of experts over those of lay subjects, as experts' intuitive judgements are laden with our best current linguistic theories while the intuitive judgements of lay subjects

are laden with folk linguistics.[4] To Devitt's critics, this is a problem for the Modest Explanation. On their view, we should not expect experts to be better subjects than lay people, since expertise in linguistics is not likely to improve one's linguistic competence in one's native language (see, e.g., Culbertson and Gross 2009 and Rey forthcoming). They also reject the idea that the linguistic intuitive judgements that are used as evidence for grammatical theories are ones that are based on subjects' theoretical beliefs (folk or otherwise) about grammaticality.

I agree with the critique of both VoC and the Modest Explanation. VoC lacks independent cognitive plausibility, and the Modest Explanation accounts for judgements which are based on theoretical concepts, which are not the kind of judgements that should be used as evidence in linguistics. Rey's view relies on special input from the speaker's competence in the form of structural descriptions. This is a substantial assumption, and one which I am not sure is necessary. In my opinion, both Rey's and Devitt's views have their benefits as well. Rey's view has the advantage that speakers' intuitive judgements are not predicted to improve with their theoretical knowledge of linguistics as their intuitive judgements are quite closely based on their linguistic competence. The Modest Explanation, on the other hand, has the advantage of requiring no special input from speakers' competence.

In the remaining parts of the chapter, I present and defend a novel account which incorporates insights from both sides of the debate. On the one hand, I want to preserve the idea that, when everything goes right, the verdict expressed in an intuitive judgement is largely due to the speaker's competence. On the other hand, I want to acknowledge the role that the central processor must play in coming up with intuitive judgements if we are to avoid the implausibility of accounts like VoC. Furthermore, I will try and accomplish this without invoking any special input from the speaker's competence. To achieve this, I need a plausible story about some signal that is relevantly related to the competence and which is accessible to the central processor. Furthermore, the story needs to explain how the central processor can fairly directly translate the signal into the content of an intuitive judgement.

The story will be roughly this: When asked to judge a sentence, the speaker will automatically (attempt to) comprehend it. The speaker's competence is involved in this step. The speaker's immediate experience of comprehending the sentence is, like on Devitt's Modest Explanation, the datum for the speaker's judgement. I point to Luka's (2005) proposal of affective evaluations as one potential way in which this experience might be available to the central proces-

[4]Though recall that Devitt argues that we should rather rely on linguistic usage data than intuitive judgements.

sor. The central processor then takes over to make the judgement. Instead of applying a (folk) theoretical concept, the central processor makes its judgement according to a non-conceptual distinction that we can be shown to be sensitive to, namely whether a sentence is part of our language or not. This second part of the account is rather like Rey's view. This distinction is due to speakers' accumulated experiences of comprehending and reacting to sentences in the past, which in turn is due to their linguistic competence, at least in part.

In the rest of this chapter, I develop this sketch in more detail. In the next section, I lay out some preliminaries of the account.

8.3 The proposed account

8.3.1 Preliminaries

Everyone in this debate agrees that "you can't help hearing an utterance of a sentence (in a language you know) as an utterance of a sentence" (Fodor 1983, quoted by Devitt 2013). Yet, as Devitt (2013) points out, "[h]earing an utterance in a certain way is one thing, judging that it has certain properties, another."[5]

I want to acknowledge that "hearing" does not automatically lead to judging. Still, I want to argue that our ability to hear an utterance in a certain way can play the central role in forming linguistic intuitive judgements. Arguably, this ability that all native speakers have of automatically "hearing" a sentence as a sentence of their language is what causes us to react to sentences that are *not* part of our language, e.g., by looking quizzically at the speaker or asking for a repetition. That "hearing" sentences as part of our language (or not) can lead to this kind of behaviour seems uncontroversial to me. However, a point which I believe is overlooked in the debate is that this ability plausibly figures in certain meta-linguistic contexts as well. The same mechanism might plausibly, I believe, be what makes us correct ourselves or others, mid-sentence or in the speaker's next turn, when something that was said was perceived to be ill-formed (this is usually referred to as "repair" in the conversation analysis literature). Below are two examples from Fox et al. (1996) from non-scripted conversations, the first from English and the second from Japanese. In the examples, we see speakers repairing their own production (note that Fox et al. use asterisks (*) to note where repair occurs):

(7) Like they-the biggest, debate ih-* in our department, in:,* at Trenton was that when we had these faculty meetings.

[5]In the remainder of the chapter, I will refer to this phenomenon simply as "hearing" or as "hearing a sentence a certain way".

(8) ja nanji goro ni kurida[shi-*]soo?
then what.time about OBL go.out
'Then about what time (shall we) go out?'

In these examples, we see the speakers reacting to their own production which they presumably perceived to be ill-formed by replacing a problematic preposition or suffix with the intended form.

In cases like these, I take it that "hearing" a sentence a certain way can give rise to conscious, verbal, meta-linguistic reactions, like corrections/repair. These cases, I believe, show that in at least some instances, the experience of "hearing" a sentence a certain way is accessible to and can serve as input for conscious mental processes. Importantly, the sources for these reactions are not generally consciously accessible to speakers. We might be able to report hearing something as being off, but not to say what caused us to pick up on it.

One might worry whether these conscious reactions are themselves the results of meta-linguistic judgements. If that is the case, then behaviour such as corrections does not support the argument I am trying to make. So we want to figure out whether the cases mentioned above are themselves plausibly based on meta-linguistic judgements.

This is an interesting question, and I do not know of any studies attacking it head on. It might be illuminated by looking at a combination of interaction data and judgement experiments designed to test whether speakers' judgements and their behaviour come apart. First, interaction data can show us what speakers normally do when confronted with an ill-formed sentence (and how that differs from their behaviour when presented with a sentence with surprising content, say). Reactions may include looking at the speaker a certain way, staying silent when it is otherwise their turn to speak, providing a correction by producing a new utterance with an altered version of the ill-formed structure, etc. (see Schegloff et al. 1977 for a classic conversation analysis treatment of repair). There is an extensive literature on this, and from that we get a way of telling that a speaker has perceived an utterance as ill-formed, which we can then use in the next step. One prominent result in this literature is that self-repair can often be shown in conversations to be preferred (treated as less problematic in the conversation) to others making corrections (Schegloff et al., 1977).

Second, if corrections are themselves based on meta-linguistic judgements, we would expect that people's reactions to an ill-formed sentence would express the same stance as their elicited intuitive meta-linguistic judgements to that sentence. What we need, then, is judgement experiments set up so as to test whether people's in-conversation corrections of themselves or others and their explicit, elicited intuitive judgements come apart or not.

I am not aware of any studies that do exactly this, but Labov's well-known

study of positive "anymore" comes to mind (Labov 1975, citing Labov 1973).[6] In Labov's judgement experiment, speakers were asked about the acceptability of sentences containing the so called positive "anymore" construction as in "John is smoking a lot anymore", roughly corresponding to "John is smoking a lot these days". Labov (1975) describes how he and other researchers through observation studies had found this phenomenon in casual conversation among white speakers of all social classes in the Philadelphia area with no evidence of social stigma. In the judgement experiments, however, judgements of the phenomenon were erratic. I want to set aside for now the question of what the cause of this mismatch between judgements and production might be. The interesting point in this context is that 12 participants who rated constructions with positive "anymore" as being unacceptable were later overhead by the experimenters using the positive "anymore" construction themselves in casual interaction. In other words, this is a case of the speaker's *own production* contradicting their meta-linguistic judgement. Now, if corrections were based on the same kind of meta-linguistic judgement as explicit linguistic intuitive judgements, we would have expected speakers who judged the structure to be unacceptable to also correct themselves when producing it (recall that self-repair is preferred to repair by others).[7]

Based on these preliminary observations, I will proceed on the assumption that reactions such as in-conversation corrections are not themselves based on meta-linguistic judgements.

If speakers more or less regularly have these experiences of "hearing" a sentence a certain way, and if those experiences are, at least sometimes, consciously available as the basis for corrections, then speakers will have an accumulated body of experiences with "hearing" sentences as being a certain way. Even if speakers have never given any specific thought to these experiences, this body of experience, I believe, at least makes it possible that speakers can be made to pay attention to this distinction and to report their judgements according to it. I say more about this in section 8.3.3 below.[8]

[6] See also Hindle and Sag (1975) on positive "anymore" and Greenbaum (1976) on differences between usage and acceptability judgements.

[7] While one could argue that this example is, perhaps, not central to the interests of most generative syntacticians as it centres on non-syntactic variation, it nonetheless illustrates my point: that if corrections in conversations were based on meta-linguistic intuitive judgements, then these speakers should be expected to correct themselves. Since they do not, this points to corrections being made on some other basis than meta-linguistic intuitive judgements.

[8] In the following sections, I will argue that in many cases, meta-linguistic intuitive judgements are based on the experience of "hearing" a sentence a certain way, which is also what I believe causes corrections like self-repair. In the Labov example above, however, this cannot be the case, as the lack of self-repair and the negative judgements point in different directions. I take it that this is a case where the process of coming up with a meta-linguistic intuitive judgement goes wrong and is influenced by irrelevant factors. I discuss this possibility in section 8.4.2.

It is, however, not trivial how the "hearing" comes about or whether this is really what is responsible for linguistic intuitive judgements. I will defend an account on which it is the case that "hearing" sentences a certain way is responsible for linguistic intuitive judgements. But before developing this account further, I would like to make a few clarifications of the view so far.

First of all, it is not clear what underlying mental architecture allows this "hearing". One suggestion is that we monitor the linguistic output of ourselves and others, comparing it to predictions of what will be said. Drożdżowicz (2018) develops this account in the case of intuitive judgements about meaning. I will leave this question open and just note that no one in the debate disagrees that we have this ability of "hearing" sentences as part of our language or not.

Second, this proposal does not make any particular theoretical assumptions about what the mental grammar of speakers looks like. Whatever mental grammar is in the speaker's mind is, together with other non-grammatical factors (e.g., factors related to semantics, pragmatics, and performance), responsible for how they "hear" a sentence, and this is what, on this proposal, is the basis for the speaker's linguistic intuitive judgements.

Relatedly, the proposal is not dependent on any particular stance on whether the structure rules of one's language are themselves implemented in the speaker's mind, or whether the processing rules that are implemented in the mind just respect those structure rules (see Devitt 2006b, ch. 3, for a discussion). Another way to phrase the latter option is that the structure rules that linguists strive to identify are specifications of the problem to be solved at the computational level of Marr's (1982) three levels of analysis rather than descriptions of what the rules actually look like on the algorithmic level.[9] On the view that I propose, all we need to be committed to is that whatever the rules/principles/mechanisms that *are* implemented in the minds of speakers look like, they cause us to hear things *as if* language was governed by the structure rules that linguists strive to uncover. However, since the view does not require any specific set of rules to be implemented in the minds of speakers, the position does not lend any particular support to views that require specific rules to be represented or otherwise embodied in the minds of speakers (see Devitt 2006c, 484-485).

In the following, I will first lay out an account of the datum for the intuitive judgement. Then, I will turn to the question of how this datum is treated by the central processor.

[9] I am not convinced that the rules proposed by generative linguists are generally interpreted within the field to be at the algorithmic level; the distinction does not seem to be regularly discussed (though see section 2.3.6). Here I mainly want to clarify my own position.

8.3.2 The datum

Devitt's account of how the datum for linguistic intuitive judgements comes about strikes me as highly plausible. As outlined above, on Devitt's account, the speaker is presented with a sentence and then "simulates" comprehending or producing that sentence. This experience is then the datum for their judgement (Devitt, 2010b). This simulation, Devitt writes, might only be necessary in cases that are hard to judge since in easy cases, we are not aware of going through any such simulation. I think another compatible way of describing this situation is that subjects automatically, involuntarily attempt to comprehend sentences presented to them (whether in written or oral form). If comprehension succeeds, they will have an experience of ease, but if comprehension ultimately does not succeed, or if it is only reached with some mental difficulty, speakers will have an experience of difficulty or downright impossibility. That experience will then be the datum for their judgement.

The experience is, in large part, due to the speaker's linguistic competence. On the account that I am developing (in contrast to Devitt's account), it is because of this that we are justified in using intuitive judgements as evidence. More on this in section 8.3.3 below.

In what follows, I present a proposal by Luka (2005) (mentioned in chapter 4), which I see as one potential way of fleshing out how this experience could present itself to the central processor. It is meant as a demonstration that a prima facie plausible story can be told about how this experience of comprehending a sentence can lead to an intuitive judgement. Whether this particular implementation story is correct is an empirical question, however, one which I do not attempt to settle here.

Luka (2005) proposes a potential link between competence and linguistic intuitive judgements as part of her Metacognitive Attribution and Preferences in Implicit Learning model (MAPIL).[10] She builds her model on research into implicit learning in general and implicit learning of artificial grammars in particular.[11]

[10] Luka's model is referenced briefly in Maynes and Gross (2013) but apart from that it has not been taken up in the debate as far as I am aware. I have found no later publications from Luka developing the model further, though Luka and Barsalou (2005) mention some of the same ideas.

[11] One could object that learning an artificial grammar in an experiment is qualitatively different from learning one's native language (and that the resulting implicit knowledge of such an artificial grammar is different from the implicit knowledge that one has of one's native language). How this objection plays out might depend on one's theory of language acquisition. On functional/usage-based accounts, subjects acquire their native language using domain-general learning mechanisms, and acquisition is aided by the important social functions that language fulfils and the motivation that comes with that, as well as the fact that natural languages have been shaped by functional constraints. On Universal Grammar-based theories, acquisition of natural languages is aided by innate rules or principles. I will not be able to argue for one position over the other here, but suffice it to say that artificial grammar learning

Luka references a study by Gordon and Holyoak (1983). In the study, participants were first presented with a number of strings generated in accordance with an artificial grammar. Afterwards, the participants were shown novel sentences, some of which conformed to the grammar that produced the sentences the participants had already seen and some of which were "ungrammatical" according to that grammar. The participants were asked both to rate how much they liked the novel strings (on a scale from 1 to 7) and to judge whether the novel strings conformed to the grammar that produced the practice strings. The authors found that participants displayed higher ratings of liking towards the utterances that were produced by that same grammar as the practice strings. Importantly, the novel strings were not identical to the practice strings, showing that this increase in affect applies to types rather than just tokens. This happened even if participants were not informed that the test sentences they had first seen were generated by a grammar till *after* they had rated their liking for the novel sentences (their ratings of grammaticality were then subsequently collected). The authors conclude that the higher rates of liking of "grammatical" sentences were caused by subjects' implicit learning of the grammar, an unconscious, automatic process. In Luka's reading, the increase in affective evaluation is due to perceived ease of processing which is due to the participants' learning of the grammar.[12] This she takes to show that affective evaluation functions as a measure of grammar learning (i.e., competence).

Based on these results, she proposes that "affective evaluations regarding ease of sentence comprehension are one of the mechanisms by which unconscious information about language enters conscious awareness and influences the report of linguistic judgments" (Luka, 2005, 483).

On MAPIL, these affective evaluations are caused by our implicit learning of the grammar (i.e., our competence), and I believe it is plausible that they are what give rise to our conscious awareness of "hearing" sentences as being a certain way.

On Luka's model, the affective evaluation is presumed to be due to ease of processing. If the sentence is in fact part of the speaker's language, the processing will, other things being equal, be effortless, and the speaker will evaluate the sentence positively. If the sentence is not part of the speaker's language, the processing will fail, and the speaker will evaluate the sentence negatively. The proposal is that this affective evaluation of a sentence is something which the speaker, when asked to make a meta-linguistic judgement about the sentence,

experiments will be most useful if the artificial grammar used resembles natural languages as much as possible (in whatever way is relevant on one's preferred theory of acquisition).

[12]Featherston (2007, 294) proposes that intuitive judgements track "computational effort", which is similar to Luka's proposal that perceived ease of processing plays a role for intuitive judgements through affective evaluations.

can consciously access and, e.g., report or base further deliberations upon.

Crucially for our purposes, such affective evaluations can be accessed by the central processor and be reported. Note that this view does not entail that we have introspective access to the *source* of the affective evaluation.

MAPIL also highlights a potential source of error for linguistic intuitive judgements. On this account, the mere exposure to a structure may increase affective evaluation of that structure, even if the structure is not part of the speaker's grammar (Luka 2005, 482-483, see also Zajonc 2001 on "the mere exposure effect"). This means that one may judge a sentence positively because it has been encountered frequently rather than because it is part of one's language. In other words, exposure might influence the *judgement* that a speaker makes in a way that does not reflect the *grammar* of that speaker's language.[13] This is a potential source of error in linguistic intuitive judgements which, I believe, linguists are aware of, and Luka's model provides an account of how it works. It is widely acknowledged that non-grammatical factors can influence linguistic intuitive judgements, and so, in my opinion, it is an advantage of this proposal that it can account for how some of these influences might come about.

It should be noted that more research into the connection between grammar learning/competence, affective evaluations, and linguistic intuitive judgements is needed to further develop this proposal. However, I hope to have shown that it is, at least, a promising place to start looking for a way for the experience of hearing a sentence to be available to the central processor as the basis for a linguistic intuitive judgement.

So far, this proposal is compatible with the Modest Explanation. The affective evaluation could be the datum that the speaker bases their judgement on. At the same time, the affective evaluations might also be compatible with Textor's (2009) "linguistic seemings" account (mentioned in chapter 2) on which a sentence simply "presents itself" as well-formed or not to the speaker. All sides of the debate agree that there is some role to play for the speaker's competence. The main difference between them, I believe, comes later, at the stage of the central processor's involvement. More on this in the next section (8.3.3).

To recap, on this proposal, linguistic intuitive judgements are based on the speaker's experience of attempting to comprehend the sentence they are presented with. This experience is in large part due to the speaker's competence, although non-grammatical factors will play a role as well. This experience, I

[13]This might also help explain why longer, more complex sentences are usually judged less positively than shorter, simpler sentences even when both are perfectly well-formed. Short and simple sentence types are likely to be more frequent, and so these sentences have the added positive evaluation stemming from the mere exposure effect on top of the evaluation based on the speaker's implicit knowledge of the structure. Longer, more complex but perfectly well-formed sentences, on the other hand, are less likely to have this added benefit.

believe, is closely related to the experience of "hearing" a sentence as being a certain way as discussed in section 8.3.1. The central processor then bases a judgement on this experience. I presented a model by Luka on which the experience is available to the central processor in the form of an affective evaluation.

8.3.3 The judgement

So far, my account has been compatible with the Modest Explanation. The datum for the speaker's judgement is their experience of attempting to comprehend the sentence they are presented with. However, the speaker's experience of the sentence could influence their judgement in different ways. It could be reported more or less directly ("That sentence sounds fine", or "That sentence seems worse than the other one"), or it could serve as the input for an immediate, unreflective deliberation like on the Modest Explanation, where the speaker's competence provides a datum which is then further worked on by the central processor. I will argue that linguistic intuitive judgements, when everything goes right, are more like the direct reports of speakers' experience of the sentence than the results of immediate deliberations based on the datum of that experience. On my account, it is because the experience of "hearing" a sentence a certain way is, in large part, due to the speaker's competence that we are justified in using intuitive judgements as evidence.

In this section, I lay out what a report of the speaker's experience might look like, and then I defend that story against a couple of potential objections in section 8.4.

As mentioned, I am looking for a way in which the datum for the intuitive judgement can be fairly directly translated into an intuitive judgement by the central processor without the application of (folk) theoretical concepts. Here, taking speakers' experience of "hearing" a sentence a certain way as the datum for the judgement is helpful. It does, of course, not directly supply the informational content for a judgement about whether the sentence is well-formed or not. However, comparatively little central processor-work has to go into transforming, say, a positive experience of a sentence, perhaps in the form of an affective evaluation as suggested above, into a judgement with the content "that sounds fine".

But, when asked to report their experience, speakers will invariably be making a kind of categorisation judgement, and so they will be applying *some* concept or distinction. If I am right that the "hearing" of a sentence as being a certain way *can* be consciously accessible, then presumably this is a feeling speakers can be brought to recognise and pay attention to, even if they do not already have a specific term for it. The concept or distinction that speakers ap-

ply in such a case need not be lexicalised for speakers to be able to recognise the feeling referred to. For a similar situation where someone is brought to recognise a feeling they have experienced but which they, so far, have no lexicalised term for, think of native English-speakers who are confronted for the first time with the concept which in German is lexicalised as *Schadenfreude* (roughly meaning the joy one gets from observing others' misfortune). Even though English lacks its own lexical item covering this particular feeling, native speakers of English who have never encountered the term before will surely recognise the *feeling* it refers to when the term is introduced and explained to them. Similarly, I believe, speakers of a language can be brought to recognise the feeling of "hearing" a sentence a certain way without having a lexicalised term for that feeling.

Let us dwell a bit on the concept or distinction applied by the speaker when making their intuitive judgement. Rey (forthcoming) presents some ideas on this issue which I find helpful. In general, having a concept of X let us reason about Xs in relevant ways. Rey (forthcoming) points to the area of visual perception, where even speakers who do not have the concept of SQUARE or DIAMOND, or even SYMMETRY would likely be able to make reliable distinctions between two images, one of a square and one of a corresponding diamond (the same square rotated 90 degrees). We want to acknowledge that they are somehow sensitive to this distinction, even though they may not have a lexicalised concept, or even a non-lexicalised concept, for the two figures which they can use for reasoning about them in the relevant ways that someone with the concept SQUARE can reason about squares. One might add that if presented with this task a number of times, it is likely that subjects will develop a concept, although perhaps not a lexicalised one, for the figures they are presented with. But the crucial point is that they do not need to have this concept ready in order to make the distinction in their first trial. Rey proposes that native speakers of a language stand in the same kind of relationship to the grammar of their language – they will be sensitive to the distinctions of the grammar in their behaviour, but they will most likely not have lexical (or even non-lexicalised) concepts corresponding to these distinctions that they can reason about.

I want to draw on Rey's account here and propose that in the case of linguistic intuitive judgements, speakers only need to be sensitive to the distinction between "hearing" sentences as being part of their language and "hearing" sentences as not being part of their language. This distinction is then the one they apply when making their linguistic intuitive judgement. On this proposal, whether or not they have a lexicalised or non-lexicalised *concept* is of less importance. Of course, it is possible that speakers have the lexical term "acceptability" (which is the term from generative linguistic theory covering this), but if they are lay subjects they almost certainly will not. It is also possible that speak-

ers will have a non-lexicalised concept with content like "how I 'heard' that sentence", but maybe they do not.

So when linguists ask whether a sentence sounds "good", "natural", or "acceptable" to a speaker, it is at least possible, I would argue, that the speaker will take the linguists to refer to that feeling of "hearing" a sentence a certain way and report that in response to the linguists' request based on their accumulated experience with hearing and reacting to sentences in the past. This should be possible even if instructions differ somewhat across different trials (asking "Does the sentence sound natural/good/acceptable to you?", etc.). But of course instructions do matter on this account.

Here one might bring up Devitt's worry that "hearing" a sentence as being one way is one thing, while it is another to *judge* that the sentence is that way. Showing sensitivity to a certain distinction in some non-verbal task might not automatically translate into making the corresponding verbal judgement. While I acknowledge this worry in general, I hope to have argued in section 8.3.1 that "hearing" a sentence a certain way can sometimes inform conscious, meta-linguistic reactions, as when speakers correct themselves or others in fluent conversation. If I am right about that, then it seems plausible to me that speakers will be able to pay conscious attention to this distinction and apply it in their judgements, even if they do not have a lexicalised term for the distinction.

So, the proposal is that, when asked in the right way for a linguistic intuitive judgement on a sentence, participants, at least in some cases, recognise that this request maps onto their feeling of "hearing" a sentence a certain way (which is, in part, determined by their competence, along with other, non-grammatical factors), and they then report that experience, applying a perhaps non-conceptual distinction. Note that they may report it with a range of different lexical items ("fine", "natural", "acceptable" etc.), the important thing being that speakers express their experience of "hearing" the sentence.

The account sketched here is an attempt at telling the story of how intuitive judgements are formed when things go right. As mentioned, however, irrelevant effects may influence intuitive judgements. The mere exposure effect has been mentioned as one potential source of error, and I will briefly mention another one in section 8.5 (Gibson and Fedorenko 2013 mention a number of other plausible candidates). Because of this, I believe linguistic intuitive judgements should be elicited and analysed with the potential sources of error in mind (something, I am sure, many linguists already do, for instance following the recommendations made by proponents of the Experimental Syntax movement, see, e.g., Cowart 1997; Schütze 2006/1996; Featherston 2007). Sometimes it might be more appropriate to rely on other sources of evidence altogether or to calibrate elicited intuitive judgements with data obtained with other methods

such as elicited production, ERP, reading time measures, or corpus data.

In this section, I have drawn on Rey's proposal that, rather than apply lexicalised concepts or even non-lexicalised concepts, speakers may make their intuitive judgements based on non-conceptual distinctions they can otherwise be shown to be sensitive to. One such distinction, which at least in some speakers may be non-lexicalised or even non-conceptual, is whether a sentence is "heard" as being part of the speaker's language or not. When asked for a linguistic intuitive judgement, what speakers need to do is to recognise that they are being asked about that feeling of "hearing" a sentence a certain way and then report that experience.

But of course, it is also possible that, rather than just being more or less directly reported, the speaker's experience serves as the datum for some further inference. In the next section, I take up some objections against the proposed view, including two objections that both press this interpretation of the etiology of linguistic intuitive judgements.

8.4 Objections

8.4.1 Just the Modest Explanation?

One pressing potential objection to the account that I present is one that Devitt raises against both Rey's view and the somewhat similar views proposed by Textor (2009) and Maynes and Gross (2013) (see also Gross forthcoming). If, instead of providing the informational content of judgements, all competence does is deliver some signal to the central processor for further processing, Devitt argues, then these views are really just instances of the Modest Explanation. The same objection can be raised against the account proposed in this chapter.

There is a possible reading of the term on which "the Modest Explanation" covers all possible accounts on which the central processor plays a role. In the following, however, I will take "the Modest Explanation" to refer to the specific central processor-based proposal made by Devitt.

I will start by noting that on the Modest Explanation, different subjects may apply different concepts when making their judgements. This is part of the reason why, on the Modest Explanation, we should prefer the judgements of experts over those of lay subjects, as experts can apply the concepts of our best current linguistic theories while lay subjects apply the concepts of folk linguistics. On the Modest Explanation, this difference is a feature, not a bug. It is part of how things are when they go right. So on the Modest Explanation, the verdict expressed in the intuitive judgement is made according to the concept that the subject applies, and that concept is taken from the speaker's (folk) lin-

guistic theory (of course, on the Modest Explanation, the speaker's competence plays a role as well in coming up with the datum for the judgement).

On the account I propose, on the other hand, when all goes well, the distinction (or, perhaps, non-lexicalised concept) that speakers apply to come up with their intuitive judgement is the distinction between "hearing" a sentence as being part of their language and "hearing" a sentence as not being part of their language (as proposed in section 8.3.1 and 8.3.3, the fact that the experience, at least sometimes, is the basis for conscious, verbal, meta-linguistic reactions suggests that it is likely that speakers can be brought to turn their attention to this experience). This experience of "hearing" a sentence as being a certain way should be the same for all speakers (even though a particular sentence will, of course, be part of some languages or dialects and not of others). Like on the Modest Explanation, the datum for the judgement is the speaker's immediate experience of trying to comprehend the sentence. But contrary to the Modest Explanation, on my account the verdict expressed in the intuitive judgement is based on a non-conceptual distinction which in turn is based on the speaker's accumulated experiences of "hearing" a sentence as being a certain way. The distinction applied in the categorisation judgement is thus itself based on the speaker's competence, so to speak. This is the main difference between the view proposed here and the Modest Explanation.

Still, on this view, there is a role to play for the central processor. "Hearing" a sentence as off is not the same as judging that it is unacceptable. But with affective evaluations accompanying sentence comprehension (or some similar signal), that gap is shortened. If the affective evaluation is actually due to the speaker's implicit knowledge of the relevant language (and not, e.g., the mere exposure effect as mentioned above), and if the instructions the subject is given succeed in eliciting a report of the affective evaluation, then "hearing" may lead, fairly directly, to judging (as sketched in section 8.3.3 above). And it is in the reporting of the affective evaluation, as described above, that the central processor plays a part.

But surely, the central processor can be involved to a smaller or larger degree. In the case of ordinary perceptual judgements of the type "the grass is brown", which is covered by the Modest Explanation, there is some central processor involvement, although we would not in general expect performance in this task to vary with expertise. This type of judgement strikes me as different, if not in kind then in degree, from the one made by an art expert who looks at a supposedly antique statue and immediately recognises it as a fake (Devitt, 2006c).

The difference between these two kinds of judgements, as far as I can see, can be construed in two different ways. On the one hand, Devitt (2012) dis-

cusses the difference between judgements made about everyday, humdrum cases and judgements made about more fanciful, hypothetical cases and argues that people's ability to make judgements about the two kinds of cases might differ (with the latter being more difficult).[14]

On the other hand, the two kinds of judgements might also differ in their degree of theory-ladenness. As Devitt (2006c, 502) writes: "On the Duhem-Quine Thesis, all judgments are theory-laden, but they are not all laden to the same degree." Devitt (2012) describes how, on the Modest Explanation, linguistic intuitive judgements are not theory-laden in the sense that they are the result of theorising. After all, most naive, native speakers of a language can produce linguistic intuitive judgements without having done any or much linguistic theorising. Rather:

> even the most straightforward judgments arising from observational experiences may depend on a background. We would not make the judgments if we did not hold certain beliefs or theories, some involving the concepts deployed in the judgments. We would not make the judgments if we did not have certain predispositions, some innate but many acquired in training, to respond selectively to experiences. (Devitt, 2012, 19)

Yet Devitt (2006c, 2012) also describes how two linguists subscribing to different linguistic theories might have different intuitive judgements as a result of them subscribing to different theories. This suggests to me that there may be a difference in degree in the way that linguistic intuitive judgements may be theory-laden: At the one end of the scale, we have judgements which are theory-laden purely in the sense that even observation judgements are, and somewhere further up along the scale we have judgements which are theory-laden in this basic sense *as well as* being theory-laden because they are made due to concepts which are taken from specific theories.

The type of linguistic intuitive judgement I am interested in on my account are the ones that do *not* involve concepts taken from specific theories, although the judgements will still be theory-laden in the basic sense in which all observation judgements are theory-laden.

To recap, the proposed view differs from the Modest Explanation in two ways: 1) The distinction (or, perhaps, non-lexicalised concept) that the central processor applies in making the intuitive judgement is due to the speaker's competence, rather than being a concept drawn from folk or expert linguistic theory. 2) The linguistic intuitive judgements that we are interested in using as

[14]This discussion is of semantic intuitive judgements in particular, but I think it extends to intuitive judgements of morphosyntactic well-formedness as well.

evidence for grammatical theories are different in degree (though not in kind) from judgements based on theoretical concepts, but they are theory-laden to the same degree that ordinary observation judgements are.

8.4.2 Acceptability or grammaticality

Another potential objection to the proposed view is based on the debate between Devitt and his critics about acceptability and grammaticality. The critics argue that we should be interested in judgements of *acceptability*, a term introduced into linguistic theoretical vocabulary by Chomsky (1965) to cover sentences that appear immediately well-formed to speakers (but might turn out not, in fact, to conform to the grammar of their language). Devitt, on the other hand, argues that we should really be interested in judgements of *grammaticality*, i.e., judgements of whether a sentence is in fact grammatical or not in the speaker's language. He also argues that lay speakers are likely to already possess the lay concept GRAMMATICALITY but not a lay concept corresponding to the linguistic theoretical concept ACCEPTABILITY. For this reason, he argues, lay subjects are likely to understand the task they are given in terms of their concept GRAMMATICALITY even if the instructions are phrased using terms like "acceptable", "good", or "natural" (Devitt 2006c, 490; see also Devitt 2010a).

This worry extends to the account proposed in this paper. Even if speakers are just reporting their affective evaluation of a sentence, they are making a categorisation judgement and so will be applying *some* concept. Since lay people (who have at least basic schooling) are likely to possess the concept GRAMMATICALITY already, should we not think that this is the concept they will apply?

The first part of my response to that is to note that the concept of grammaticality that lay subjects are likely to possess is presumably a prescriptive one since grammar is generally taught prescriptively.[15] By wording the instructions to subjects with terms like "natural" instead of "grammatical", linguists are trying to get at participants' intuitive judgements without the prescriptivist attitude.[16] This, I think, is reasonable, as speakers are sometimes able to say that a construction sounds natural to them although they know that it is not part of the prescriptivist grammar of the standard variety of their language (think of dialects or stigmatised varieties).

[15]Devitt (2010b) points to the fact that there is a large overlap between a speaker's prescribed language and their actual language. So, although a few judgements might be based on the prescribed grammar, on the whole most should be based on the speaker's experience with their actual grammar.

[16]Culbertson and Gross (2009, 734) write in the instructions to the participants in their study: "Here, we are interested in your linguistic intuitions, not in the rules of 'proper English' that you might have been taught in school."

However, Devitt (2006c, 2010a) argues that it is likely that, in spite of linguists' intentions, folk participants will give grammaticality judgements in these cases. Referring to Grice, Devitt writes that "[i]n an appropriate context, there may be nothing about an expression that the linguist could plausibly be seen as asking other than about its grammaticality" (Devitt, 2010a, 842). But I would think that if the speaker's everyday concept of grammaticality is a prescriptive one, then the linguist's instructions to give a non-prescriptive judgement might, per the Gricean maxims, lead the participant to think that something else than their everyday concept of grammaticality is called for.

If participants receive instructions that ask whether a sentence sounds "natural" or not and which explicitly excludes prescriptive senses of grammar, I think it is likely that lay participants will interpret this to exclude theoretically based theory-laden judgements infused with prescriptive folk linguistic theory, such as "it's wrong to end a sentence with a preposition".

As mentioned, we do not have access to the source of our experience of "hearing" a sentence a certain way, and so in cases where a prescriptivist attitude is very engrained in the speaker, it is likely that their experience of "hearing" a sentence a certain way is based at least partly on a prescriptivist notion of grammar.[17] Along with the mere exposure effect, this is another potential source of error that linguists using intuitive judgements as evidence should look out for (and many, undoubtedly, do).

The discussion above presupposes that speakers have a certain notion of grammaticality. This is likely to be true of speakers from WEIRD (Western, Educated, Industrialised, Rich, Democratic) backgrounds. However, this assumption is probably much less plausible in the contexts where a lot of fieldwork is carried out: on previously little-described languages that lack a written tradition. What potentially relevant concepts such speakers might bring to the table is likely to vary from culture to culture (e.g., depending on whether there is a strict division between a colloquial language and a formal language, and whether speakers might interpret the linguist's question to pertain to one rather than the other). Linguists collecting intuitive judgements will have to be sensitive to such potential influences and adjust their elicitation techniques and analyses accordingly.

I would also like to dwell on the point that just because we have a lexicalised concept which is thematically relevant to a categorisation task at hand, that does not mean that we necessarily apply that concept. Imagine that I speak a language that does not lexically distinguish mice from rats (which some lan-

[17] It is possible that this is the explanation for Labov's (1975) observation, mentioned above, that some speakers who themselves use the construction positive "anymore" deny its acceptability when asked about it.

guages do not). It is arguably plausible that I would still be able to understand instructions that ask me to pick out animals of one type but not the other, and I would likely be able to make a distinction between the two types in a majority of cases (but maybe not flawlessly). In this case, making a rat vs. mouse categorisation judgement, I would not be relying on my lexicalised concept since it does not distinguish the two but rather on some non-conceptual distinction or non-lexicalised concept, perhaps one that I cook up on the spot (see Devitt 2006c, 502-503, about learning concepts in the course of an experiment).

Similarly, since competent speakers at least sometimes can make verbal, meta-linguistic reactions based on how they "hear" sentences (as discussed in section 8.3.1 above), making this experience of "hearing" a sentence a certain way one they can be brought to consciously acknowledge, and since we are giving subjects instructions that are likely to be different from their pre-existing prescriptivist concept, I think it is plausible that speakers will not employ their regular concept of grammaticality. Instead, I laid out in section 8.3.3 how they might map their accumulated experiences of "hearing" a sentence as being a certain way onto the choices of positive and negative categories provided to them in the intuitive judgement task.

8.4.3 A version of VoC?

Since the account I present gives the speaker's competence a more prominent role than on Devitt's Modest Explanation, one could perhaps accuse it of moving dangerously close to VoC. Devitt (Devitt 2006c, and subsequent papers) raises a number of objections to VoC. I will briefly outline a few central ones here and argue that they do not apply to the account that I defend.

One objection that Devitt raises against VoC is that, if competence is so central to our intuitive judgements (if they are *the voice of competence*), why are intuitive judgements not phrased in the language of competence? More specifically, if grammaticality is the central concept of competence, and if competence delivers the fully fledged informational content of intuitive judgements, one would expect the intuitive judgements to be phrased in terms of grammaticality, not whether the sentence "sounded good", "is acceptable" etc. (Devitt, 2006c, 489). I take this objection to be based on the idea that on VoC, the speaker's competence provides the *informational content* of intuitive judgements. Since my account does not assume this, there is no reason, on my account, that we should expect competence to "use its own language" (Devitt, 2006c, 489). The same, I would argue, goes for accounts like Rey's (2013) where the informational content of intuitive judgements is not provided by the speaker's competence.[18]

[18]This objection is potentially also related to Devitt's view that the intuitive judgements

A related objection that Devitt raises against VoC is that, if competence really "speaks" through intuitive judgements, we would expect speakers to have intuitive judgements about all the different things that are part of speakers' competence (according to our current best linguistic theories, whichever those may be):

> Ordinary speakers have many intuitions about grammaticality, coreference, and ambiguity but few about transitivity, heads, A'-positions, c-command, cases, transformations, and so on. [...] If our competence consists in representing this theory and our competence speaks to us at all, how come it *says so little?* (Devitt 2006c, 489, original emphasis)

I think this objection is related to the one just mentioned above: If our competence delivered the informational content of intuitive judgements, why would it not use its own terms, and why would it not apply to all the different aspects of the grammar? (see, e.g., Fitzgerald 2010 for a reply.) I think a straightforward response to this is that, if we reject the first objection, that intuitive judgements should use the terms of competence, then we can also reject the second one. Linguistic intuitive judgements do in fact relate to many parts of the grammar, but since they do not use the terms of the grammar, it is the task of the linguist to tease out *why* a certain structure is deemed unacceptable by a speaker. A structure might be unacceptable, say, because of a violation of an island constraint, and the speaker therefore has a certain experience trying to comprehend the sentence. The speaker reports that experience. It is then up to the researcher to look at this data point in connection with other data points, current theories etc. and, if necessary, come up with more tests to try to figure out what the most likely source of the problem is. In other words, the competence does not say so little, but the fact that it does not speak in its own voice means that it is not straightforward what the source of any given problem is. On this perspective, "(un)acceptability" is a cover-term for all the different things that can go right or wrong with sentences. The intuitive judgement that a sentence is (un)acceptable may be caused by any number of factors, but the judgement will not specify which factor underlies the judgement. This is in line with the aspect of my proposal that we do not have introspective access to the source of our intuitive judgements.

Finally, Devitt argues that VoC is immodest for resting on the assumption that the rules of a speaker's grammar are represented in their mind. Instead, he argues, the Modest Explanation is preferable in that it "make[s] do with cognitive

that should be used as evidence are judgements *about* grammaticality (see sections 2.3.2 and 2.3.4).

states and processes that we are already committed to" (Devitt, 2006c, 505). Hopefully the sections above have shown that the same can be said for my proposal, and so this objection does not apply to the account I propose.

8.4.4 Special input: Rey's evidence

In his chapter on linguistic intuitive judgements, Rey (forthcoming) presents an array of empirical evidence that he finds provides a "strong prima facie case" for his view that non-conceptual structural descriptions inform the central processor when coming up with intuitive judgements – and against Devitt's view. In this section, I review the evidence and argue that none of it is incompatible with the account I have presented here.[19]

8.4.4.1 Involuntary parsing

The first point Rey raises is involuntary parsing. It is the point raised in section 8.3.1 above that we cannot help hearing a sentence uttered in a language that we know as a sentence of that language. Devitt (2013) has argued that the reason we hear a sentence as having certain properties is because our mental representation of that sentence itself *has* those properties, not because it explicitly *represents* those properties. Rey (forthcoming) points to the fact that the mental input to the central processor will have a vast number of properties, including "multitudes of physio-chemical properties" and relational properties, such as "being n minutes older than y", and only some of the input's properties are going to be relevant. How is the central processor to choose what is relevant?

Rey (forthcoming) finds that "[i]t's hard to see what else could select among the infinitude of properties a state has just the ones the hearer 'hears' the representation 'as' possessing, other than computation over representations of them." On his view, the non-conceptual structural descriptions are representations of the properties of the sentence, which can then be accessed by the central processor.

Rey (forthcoming) notes that "there's a general explanatory problem in accounting for how anything in a universe of 'local' physical interactions can in

[19]Furthermore, one might argue that the evidence that Rey presents all shows that speakers are sensitive to certain distinctions in their *behaviour*. This, in itself, does not mean that the same distinctions will be available to the speaker to make *intuitive judgements* based on. This is a central point in Devitt's challenge against his critics' accounts ("hearing" is not the same as judging). Rey notes this and writes: "I shall be assuming that if some conscious perceptual task is sensitive to certain phenomena, that is a *prima facie* reason to suppose that phenomena is in some way (i.e., at least non-conceptually) available for intuitive verdicts." He thus seems to assume the connection that Devitt is urging him to argue for. I shall not pursue this issue further here, except by noting that my account is an attempt at bridging the gap between "hearing" and judging by proposing that the connection lies in the speaker's (consciously accessible) experience of "hearing" the sentence a certain way.

general be sensitive to non-local, relational or non-physical (what I call 'abstruse') properties". One solution to this problem is that the non-local properties are represented. However, Rey also mentions the possibility that the central processor might work on "surrogate" local properties (the way a person can be identified by their fingerprint, even though the fingerprint does not, as such, "represent" that person). This is the route taken by the account I have presented here. On this account, the central processor receives the speaker's experience, possibly in the form of an affective evaluation, and it is this local, surrogate property which is the basis for the speaker's judgement.

8.4.4.2 Sensitivity to register

Another point that Rey brings up in favour of his account that non-conceptual structural descriptions inform the central processor when coming up with intuitive judgements is that children can be shown to prefer a prestigious register over the register that they are surrounded by daily. He references a study by Kinzler et al. (2012) which tested the language preferences of a group of South African children aged 5-11. The children's first language was Xhosa, but among those who attended an English language school the researchers found a preference for English over Xhosa. The very fact that the children are sensitive to this difference in social prestige between the two languages shows, according to Rey, that this is a distinction their mind is already making (presumably, on Rey's view, based on representations of register features). Rey takes this to favour his account over Devitt's, on which intuitive judgements are not based on any special input from the speaker's competence. The fact that these children may only later be able to express their meta-linguistic judgements about the matter, Rey takes to show that "[i]t likely takes a while, and specific motivations and cultural practices, to ordinarily become sensitive to one's own sensitivities, i.e., attentively enhancing and maybe giving word to distinctions your mind's been making all along".

In this case, I think everyone would agree that the sensitivity shown by the children show us that their minds are making this distinction. However, as Devitt stresses, one cannot freely assume that "hearing" leads directly to judging, and so Devitt might even take this evidence as a point in his favour.

What my account can say about this is similar to what Rey says himself. Even though we can be shown to be sensitive to something, it might take some time before we have self-awareness enough to direct our attention to our own experiences. In this case, the children will have a different experience listening to their preferred register as opposed to listening to other registers. Once they have experienced it a number of times, and perhaps when they are old enough

to have developed the right kind of self-awareness, and potentially only when something or someone, e.g., an experimenter, directs their attention to this type of experience, will they be able to make judgements based on these experiences. In contrast to Rey's view, I propose that what the children will base their judgement on is their experience of listening to the registers, not any representations made available to the central processor.

8.4.4.3 Form over meaning

The last three points that Rey brings up I will treat under one heading: form over meaning. He first points to cases of "meaningless" syntax, such as Lewis Carroll's poem *Jabberwocky*, which speakers are able to syntactically parse even though they cannot get any distinct literal message from the sentences (and so, we can judge that the sentences of Jabberwocky are syntactically well-formed even though they are lexically meaningless). His second point relates to syntax trumping the message, for instance in garden-path sentences, where it is sometimes the case that hearers will hold on to the semantic interpretation associated with an initial syntactic parse (the garden-path one) even after hearing the full sentence which semantically is incompatible with the initial interpretation.[20] The last point that he brings up is the parsing of null elements, which also points to the fact that the way subjects hear a sentence may be different from the message of that sentence.

Rey uses these points to argue against what he takes to be a component of Devitt's position: that what is available to the central processor is just the *message* of the sentence heard, and that this is the datum that linguistic intuitive judgements are based on (he points to a quote where Devitt 2006b, 112, lays out the task of the language module in comprehending sentences). I am not quite sure whether this construal of Devitt's view is accurate (if nothing else, Devitt, as mentioned, talks about the representation delivered to the central processor *having* certain properties, so at the very least the central processor also seems to be sensitive to the form of the message, on his view). But even if I am wrong about Devitt's view, it is not *my* position that the central processor only works on the message of the sentence. Instead, on my account, the central processor works on the speaker's experience of trying to comprehend the sentence (which will include, if the comprehension succeeded, the meaning of the sentence). On my view, this experience is shaped in part by the syntax of the sentence

[20]The authors of one of the studies he cites seem to be slightly more cautious, however. Ferreira et al. (2001) write: "this misinterpretation effect is somewhat selective. It does not survive if the misinterpretation is implausible". The authors also suggest that "people are often satisfied with inaccurate representations based on incomplete processing of the sentence", meaning that subjects might not have *noticed* that the full sentences is semantically incompatible with their initial interpretation.

(although, in contrast to Rey's view, no representation of the syntax of the sentence is available to the central processor). For this reason, I take it that these points do not apply against my view.

8.4.5 Other judgements

In this chapter, and in this whole dissertation, I have focused on intuitive judgements of morphosyntactic well-formedness. However, linguists use other intuitive judgements as evidence for their theories as well. Even within the study of syntax, judgements of other types are used, such as judgements of co-reference (e.g., by asking speakers things like, "in the sentence 'Agnes asks why her sister is slapping herself', can 'herself' be Agnes?"). Outside of syntax, linguists use intuitive judgements within phonology, semantics, and pragmatics as well (Maynes and Gross, 2013). Can my proposal account for these other types of judgements?

In the speaker's experience of comprehending a sentence, one component will be the meaning they have extracted from it (if the comprehension succeeded), even if they have comprehended it differently than intended by the person who produced the sentence (in cases of misunderstanding). When hearing a sentence containing co-reference, part of the experience will therefore be an understanding of who is doing what to whom (this understanding will, at least partly, be due to the syntax of the sentence). Based on this understanding, speakers will be able to answer questions about whether, e.g., "Agnes" and "herself" can co-refer or not.

As for judgements about other areas of linguistics, I believe it might be possible to tell a story similar to the one I have told here. The speaker attempts to comprehend the utterance, and this experience will include (automatic, involuntary) phonetic/phonological, semantic, and pragmatic analysis. Based on this experience, the speaker may then make intuitive judgements. Like with judgements of morphosyntactic well-formedness, the proposal is that speakers will have prior experiences of "hearing" an utterance as being either "off" or not in the relevant ways (phonological/semantic/pragmatic), and that with the right instructions linguists can make them pay attention to this experience when eliciting intuitive judgements.

8.5 Expertise

Finally, I want to consider the discussion about whether to use experts or lay subjects when eliciting linguistic intuitive judgements to use as evidence for grammatical theories. It follows from Devitt's account that linguists with good

theories of grammar will supply intuitive judgements that are better evidence for the actual grammar than the intuitive judgements made by lay subjects who are making their judgements based on folk theories of grammar. This has been called the expertise defence. [21] But recall that this was also the main point of criticism against his view. On his critics' views, experts do not have any special advantage over lay subjects since we are all, pathologies aside, equally competent speakers of our native languages.

One can imagine two things that would make the judgements of experts better than those of lay subjects:[22] their *experience* making the relevant kind of judgement, and their possession of refined *theories* and/or *concepts* to apply when making the judgements.[23]

Let us take Devitt's example of an art expert who looks at a statue and intuitively judges it to be a fake. The expert might draw on previous experience with making this kind of judgement as well as on highly specialised concepts and theories (e.g., about carving techniques, stone quality, etc.) that might not be commonly known among lay subjects. Either one of these, or a combination of both, would likely make the expert's judgement better than that of a lay person. If we take the case of linguistics, linguists might have more practice with making the relevant intuitive judgements, and they are likely to have a more refined concept and theory of grammaticality to apply than lay subjects do. So why, on the account proposed here, does this not lead to experts having better linguistic intuitive judgements than lay subjects? I will start with the issue of theories and concepts and then come back to experience.

Linguists almost certainly have more specialised theories and concepts relating to grammaticality than lay subjects do, and if the relevant judgements are formed by subjects applying their theoretical concept of grammaticality to a sentence, then linguists plausibly have an advantage over lay subjects similar to the one that the art expert has in the example above. However, if my reply to the first objection considered above (in section 8.4.1) was successful, then that kind of judgement is not the relevant one in the case of the linguistic intuitive judgements under consideration. The relevant kind of judgement is instead one

[21] The argument that experts' intuitive judgements make for better evidence than those of lay subjects is made within meta-philosophy and the Experimental Philosophy debate as well (see, e.g., Nado 2014a for a discussion).

[22] The way in which experts' judgements would be "better" does not matter here. Maybe they are able to judge correctly in a larger number of cases, or maybe they are able to make correct judgements about more detailed matters, but I will set that question aside.

[23] The first option roughly corresponds to the proposal of Weinberg et al. (2010) that philosophers' expertise with intuitive judgements could plausibly consist in practical know-how about considering hypotheticals. The second option is roughly parallel to their suggestion that this expertise could come down to the "evaluative component of the intuiting process – the stuff that, once you're entertaining the hypothetical, tells you whether or not you've got an instance of the target property on your hand" (Weinberg et al., 2010, 346), either through better conceptual schemata ("concepts") or mastery of entrenched theories.

where the subject (expert or lay person) reports their experience of "hearing" a sentence a certain way (using some available evaluative term). Thus, on this account, linguists' having refined theories or concepts does not offer support for the expertise defence.

How about experience then? Of course, as Devitt points out, we all have everyday experience of judging sentences we hear, and this might be the reason that lay subjects are relatively reliable judges about sentences. Still, linguists (at least those with the relevant specialisations) might have even more experience with making these judgements, and that might make their judgements even better evidence than those of the already pretty reliable lay subjects.

And maybe there is a sense, even on the account proposed here, in which the experience that experts have with making this type of judgements gives them an advantage over lay subjects. Experts are more likely to know in advance what the task at hand is, and so they might be less likely to misunderstand the instructions, for instance by not reporting their experience of "hearing" the sentence as being a certain way but instead applying their folk prescriptivist notion of "grammatically correct" or their favourite grammatical theory (or other, fundamentally theoretical assumptions they might have about language, such as requirements of parsimony etc.). This seems to me to be a more minor advantage than the one argued for in the expertise defence (at least, it does not seem to warrant the use of experts over properly instructed lay subjects), but even so it is an advantage. So in that sense, there is some room for experts to perform better than lay subjects on this account. This advantage can likely be counteracted by phrasing instructions to lay subjects more clearly as well as by developing more sophisticated test materials, however.[24] Furthermore, this advantage also does not counteract other sources of bias that the experts (as well as lay subjects) are susceptible to.

Theoretical bias is often mentioned as an issue that is more critical when using expert subjects compared to using lay subjects (e.g., Wasow and Arnold 2005, see also chapter 3). Another potential bias is the mere exposure effect as mentioned in section 8.3.2. I would argue that this effect is potentially more problematic for experts than for lay subjects. Linguistic researchers are much more likely than lay subjects to have been exposed to linguistic structures which might skew their judgements. If a researcher is interested in a particular construction and sets out to study it using intuitive judgements, the sheer exposure

[24]Note that this is not necessarily a question of giving lay subjects explicit "knowledge that" about how to perform the task but rather about designing materials in such a way that subjects will be responding to the issue of interest, whether they explicitly know that this is what they are doing or not. Linguistic experiments performed with children subjects provide nice examples of how test materials and the experimental set-up can compensate for subjects' lack of explicit "knowledge that" about what the task is.

to a large number of sentences of that structure could make it seem more acceptable than other constructions which are equally good but which simply have not been encountered as frequently. Furthermore, since the mere exposure effect generalises across tokens to types, it is also possible that a researcher's exposure to data from languages or dialects different from their own could influence their judgements on their native variety.[25] Similarly, lack of contact with other native speakers of one's native language might also skew one's linguistic intuitive judgements.

For the expertise defence to kick in here, we would need reason to believe that expertise will make you be less susceptible to the mere exposure effect than lay subjects are (but see Weinberg et al. 2010 and Devitt 2012 for a discussion of whether this is required on the expertise defence). Maybe something in the training or practice of linguists could have just that effect. In other sciences, practices to avoid biases, such as double blinding, are deliberately put in place and taught to aspiring researchers. However, as far as I know, no such formal training is in place in linguistics programs to train aspiring linguists in avoiding the mere exposure effect as it applies to linguistic intuitive judgements. Similarly, in linguistics papers that use the author's own judgements as evidence, one very rarely gets a report of the order of presentation of examples and other factors which might be relevant for the mere exposure effect.[26] At best, individual linguists might have ad-hoc, home-made procedures for dealing with the mere exposure effect. Based on this, I find it unlikely that, across the board, expertise in (relevant sub-fields of) linguistics makes linguists less susceptible than lay subjects to the mere exposure effect.

Summing up, experts may have a practical advantage over lay subject in knowing the task already. This practical advantage can, however, likely be balanced out by using the right instructions and test materials. On the issue of the mere exposure effect, we have seen that experts may in fact be *more* likely than lay subjects to be in a situation where they are affected by this irrelevant effect. All in all, in contrast to Devitt's Modest Explanation, this account does not provide support for the expertise defence regarding intuitive judgements in linguistics.

8.6 Conclusion

In conclusion, the answer that this account offers to the justification question is that linguistic intuitive judgements can be used as evidence for grammatical

[25]Anecdotally, I experience that after speaking Swedish for a while, my linguistic intuitive judgements about Danish, my native language, are skewed.

[26]Though the Experimental Syntax movement (chapter 3) wants that to change.

theories because they are reports of the speaker's immediate experience of trying to comprehend the target sentence. The justification thus ultimately comes from the speaker's competence, which is, at least in part, responsible for the speaker's experience of the sentence (when all goes well). The account fills the gap between on the one hand Devitt's Modest Explanation and on the other hand Rey's competence-based view by not requiring any special input from competence while also not employing any (folk) theoretical concepts in the judgement process. As the speaker's experience has to be reported to figure as evidence, the central processor has a part to play as well in the story, and on my account, this leaves linguistic intuitive judgements theory-laden to the same (relatively low) degree that observation judgements in general are. With this proposal, I hope to have shown that there is space for an account of the etiology of linguistic intuitive judgements that allows the speaker's competence to play a central role in the formation of judgements without relying on cognitive assumptions that lack independent plausibility and without relying on special input from the speaker's competence.

As for the expertise defence, this account does not provide support for arguing that experts' intuitive judgements make for better evidence than those of lay subjects. This is with the exception of the fairly trivial point that experts might be more convenient to use as subjects since they will potentially need less specific instructions and/or sophisticated test materials as they already know what the task is. This is in contrast with Devitt's Modest Explanation, on which we should expect experts to perform better than lay subjects.

This account also points to some potential pitfalls of using intuitive judgements as evidence in linguistics. Since we do not have introspective access to the source of our experiences of sentences, and since they can be impacted by other things than linguistic competence (like, for instance, by the mere exposure effect), we need to (continue to) take relevant precautions when using this type of evidence and in some cases look for other sources of evidence.

Chapter 9

Conclusions and closing notes

In this dissertation, I investigate the theoretical assumptions underlying linguists' use of intuitive judgements about the morphosyntactic well-formedness of sentences as evidence for grammatical theories. Furthermore, I present and defend a novel account of the etiology of linguistic intuitive judgements, one that, if roughly true, can justify their evidential use, at least under certain circumstances.

In this concluding chapter, I first sum up and comment on the results of the empirical study described in chapters 5 to 7 and how those results relate to the etiology debate. In the next part of the conclusion, I turn to the account of the etiology of linguistic intuitive judgements that I developed and defended in chapter 8 and discuss how it answers the justification question. Then I return to some of the wider debates that are relevant to the central theme of this dissertation, namely the debate over the evidential use of intuitive judgements in philosophy and the theoretical divide within the field of linguistics. Finally, I sum up this dissertation's contribution to the debate over the etiology of linguistic intuitive judgements and the debate on the justification for using them as evidence for grammatical theories.

9.1 Empirical results

The dissertation takes as its starting point the debate about how the etiology of intuitive judgements might justify their evidential use, as well as the debate over how to best collect and analyse intuitive judgements when using them as evidence in grammatical research.

Until now, the beliefs of linguists about the justification for using intuitive judgements as evidence for grammatical theories have been left largely undiscussed within linguistics, and while the issue has been at the centre of a debate

within the philosophy of linguistics, the voices of practising linguists have largely been missing from that debate.

Through a questionnaire study, I show that adherents of different theoretical orientations in linguistics have different assumptions about intuitive judgements as evidence, specifically comparing generative and non-generative linguists. On the generative side, the majority view is that intuitive judgements make good evidence for grammatical theories because they express the speaker's linguistic competence, and that, if one could filter out the noise from performance factors, intuitive judgements would be infallible about the grammatical status of sentences in the speaker's language. Among the non-generative participants, the majority view is that intuitive judgements make good evidence for grammatical theories because they express the speaker's reflections about their language. Non-generative linguists do not think that intuitive judgements would be infallible if we could filter out noise from performance factors.

The majority of both generative and non-generative participants reject the idea that the content of intuitive judgements is deduced from the rules of the speaker's mental grammar and furthermore reject the idea that the structure rules of a speaker's language need even be implemented in the speaker's mind.

A majority of the non-generative participants think that only acceptability judgements can be used as evidence in linguistics (as compared to grammaticality judgements), and they generally subscribe to a non-mentalist conception of the aim of grammatical research. There is no clear majority on these two issues among the generative participants.

As for the Experimental Syntax debate, there is no clear majority among the generative participants for either the traditional armchair method of collecting and analysing intuitive judgements, nor for the experimental method of using large samples of materials and subjects, analysing the results with appropriate statistical methods. Among the non-generative participants, on the other hand, the majority view aligns with the experimental side of that debate.

These findings are, I believe, interesting in themselves for the areas of linguistic methodology and the philosophy of linguistics. However, they are also relevant to the debate between Michael Devitt and a number of philosophers of language and linguistics about what the received view on this question is within generative linguistics. The view that I found to be the majority view among generative linguists does not match the Voice of Competence view (VoC) that Devitt attributes to generative linguists (of which it is a central aspect that the speaker's linguistic competence provides the content of their intuitive judgements). However, it does not match the view defended by Devitt's critics either, nor the view that Devitt himself argues for. So while I have shown that a majority of generative linguists do not seem to subscribe to VoC, they do not

subscribe to any of the other well-developed views found in the debate either. While this result is perhaps not surprising, this study is the first to answer the empirical question of what the majority view on the etiology question is among generative linguists. Developing a coherent etiology story and accompanying justification for the evidential use of linguistic intuitive judgements on the basis of the majority answers from the generative participants is a task that is up to future work by anyone who wants to defend that position. Interestingly, the majority view of the non-generative participants was fairly close to the Modest Explanation argued for by Devitt.

Association analyses show, however, that generative participants are significantly more likely than non-generative participants to hold a number of the views associated with VoC. Generative linguists are more likely than their non-generative colleagues to think that intuitive judgements are good evidence for theories of grammar because they reflect speakers' competence, and that intuitive judgements are actually deduced from the speaker's mental grammar. They are also more likely than non-generative participants to think that intuitive judgements would be infallible if performance factors could be filtered out, that it is a good hypothesis that structure rules are implemented in the minds of speakers, and that the ultimate subject matter of grammatical research is the mental capacities that underlie speakers' linguistic abilities.

In short, the association analyses show that on almost all the issues of the etiology debate, the generative linguists are more likely than their non-generative colleagues to embrace VoC. This might be one reason that VoC seems to be prominent in the generative literature, despite not in fact being the majority generative view: It involves a lot of areas where generative and non-generative views clash. Still, in relation to the etiology debate, this result does not mean that a majority of generative linguists embrace VoC, even if they are in general more likely to do so than non-generative linguists.

9.2 My justification account

In the dissertation, I also develop and defend a novel account of the etiology of linguistic intuitive judgements which draws on insights from the two main competitors in the etiology debate. On the one hand, my aim was to develop an account that does not include any special input from the speaker's competence and which does not rely on cognitive assumptions that lack independent plausibility (as VoC does). On the other hand, I wanted an account that does not predict that speakers' knowledge of formal linguistic theory influences their intuitive judgements when all goes well. The kind of judgement that is based on subjects' knowledge of theories, is, I believe, simply not the one linguists

are after when they collect intuitive judgements to use as evidence for their theories. If the best available account of these intuitive judgements turns out to be one on which this is in fact the kind of judgements that linguists employ, I believe that linguists should cease to use this kind of judgements as evidence for theories of grammar (Devitt similarly argues that they should give way to evidence from usage). They might still, however, be used for coming up with hypotheses or talking informally about cases that are clear to everyone (e.g., "the cat is on the mat" is well-formed, see Santana forthcoming for an account of how intuitive judgements might be used as an appeal to common ground rather than as evidence as such).

However, if linguists want to use intuitive judgements as *evidence* for their theories, we need a plausible justification for that practice. What story about intuitive judgements could one tell that would justify their use as evidence?

With my account, I wanted to expand on how the speaker's linguistic competence could play a role, when all goes well, short of supplying the content of intuitive judgements (as it does on VoC). On the account that I develop, intuitive judgements are based on the speaker's immediate experience of hearing and automatically trying to comprehend the sentence. I suggest that this experience might be what causes speakers to correct themselves or other speakers in conversation when perceiving something that was said to be ill-formed. I reviewed Luka's (2005) proposal that intuitive judgements are based on a feeling of affective evaluation towards the sentence, which is caused by the speaker's implicit "knowledge" of the grammar. This is one concrete suggestion for how the experience of hearing the sentence a certain way becomes available to the speaker's central processor, but more research is needed to investigate that further.

So far, my account is in principle compatible with Devitt's Modest Explanation. Where I part from that account is on the question of how the central processor arrives at the judgement. I acknowledge that these judgements will be theory-laden to the degree that all observation judgements are theory-laden. However, I do not agree with the Modest Explanation that the concept applied by the speaker's central processor has to be the speaker's concept of grammaticality for the intuitive judgements to be interesting as evidence for grammatical theories.

Instead, I propose that the speaker will be able to make a non-conceptual distinction based on their accumulated experience with individual experiences of "hearing" a sentence in a language that they understand as a sentence of their language (or not). I argue that, if I am right that this experience is what makes speakers correct themselves and other in conversation, then this feeling is at least sometimes available as the basis for verbal linguistic reactions. I propose that, in that case, speakers can presumably be made to attend to that

distinction, even if they do not have a lexical term corresponding exactly to that experience.

On this account, there is no special input from the speaker's competence to their central processor and no direct access to truths about the speaker's language through their intuitive judgements. Linguistic intuitive judgements are ordinary judgements based on the speaker's immediate experience of hearing the sentence in question and their accumulated previous experiences of hearing and evaluating sentences in this way. This entails, as mentioned, that intuitive judgements are theory-laden, at least to the degree that all observation judgements are. Since we do not have introspective access to the sources of our judgements, on this account intuitive judgements are also open to influences from other sources as well. If the speaker has very ingrained prescriptive attitudes, this might present itself as a strong reaction against a sentence that is, in fact, part of the speaker's language. It is also possible that the written language bias that Dąbrowska (2016) describes can be accounted for in this way. Another potential source that might influence speakers' judgements is the mere exposure effect mentioned in chapter 8. It is not transparent to speakers what the sources of their intuitive judgements are.

The account is not based on the results from the questionnaire, detailing the majority views within generative and non-generative linguistics respectively. Still, one might ask whether it is compatible with either or both of those majority views? On the account that I propose, linguistic intuitive judgements are fairly direct reports based on speakers' linguistic competence, but those reports have to be mediated by central processor reflection based on previous experiences. According to my account, other things than purely performance factors in the traditional sense can influence intuitive judgements (ingrained prescriptive attitudes, for instance), so intuitive judgements would not be infallible even with performance factors filtered out. This means that the account is in principle compatible with the non-generative majority view, on which intuitive judgements are due to speakers' reflection on their language, and on which intuitive judgements would still be fallible even if performance factors were filtered out. Whether non-generative linguists would in fact endorse this account is, of course, another question. The account is not so immediately compatible with the majority generative view, specifically because on that view, linguistic intuitive judgements would be infallible if performance factors could be filtered out.

However, while the account might not be in line with what turned out to be the majority generative view on the etiology of linguistic intuitive judgements, I do not think the account is incompatible with generative principles in general. For instance, competence still plays a central role on the account, although it

is mediated by central processor reflection. The account does not rely on any particularly contentious assumptions about language, as far as I can see, and so in principle, it should be compatible with both generative, non-generative, and so called atheoretical or descriptive approaches to linguistics.

My proposal is an attempt to provide an account for the etiology of linguistic intuitive judgements which may justify their use as evidence for grammatical theories without relying on cognitive assumptions that lack independent plausibility and without entailing that speakers' knowledge of linguistic theories influences their intuitive judgements, at least, when all goes well. However, whether this account is ultimately the *empirically correct* account of how the linguistic intuitive judgements used as evidence for grammatical theories actually come about is, of course, open to further research.

9.3 Implications

Let us turn now to some of the lessons we can draw from the conclusions presented above. I want to mention what the justification laid out above means for the practice of using intuitive judgements as evidence theories of grammar. In short, it is not terribly dramatic. On my account, linguistic intuitive judgements are generally reliable though not infallible, and so can be used as evidence for grammatical theories when proper care is taken.

However, as mentioned, the account does entail that intuitive judgements may in some cases not reflect the speaker's linguistic competence. A general reason for this is that participants may misunderstand the instructions they are given. While it is possible that this risk is minimised when using experts rather than naive subjects, this potential difference can be countered with better designed test materials and/or more precise instructions, and this advantage does not make experts less susceptible to other potential sources of error. Furthermore, this is more of a general problem for this kind of data collection than a particular problem for my account, and I will not discuss it further here.

On this account, a more severe risk when using linguistic intuitive judgements as evidence is, as mentioned above, that they will not reflect the speaker's linguistic competence if some other source influences the speaker's experience of "hearing" the sentence a certain way. For this reason, known or suspected potential sources of noise (what Fischer and Engelhardt 2016 call "vitiating circumstances") such as the mere exposure effect and other potential influences should be balanced out by relevant practices in the collection and analysis of linguistic intuitive judgements.

It is possible that some such vitiating circumstances are already balanced out by the practices of linguists using this type of evidence. However, one of

the problems with the currently widespread informal, traditional method of collecting and analysing intuitive judgements is that it is intransparent. The researcher who uses themselves as subject does not note in their paper which order sentences were considered in, whether further sentences were considered but discarded, or how many colleagues were consulted about the sentence. For us to say that a particular vitiating circumstance is countered by the practices of researchers, we need to know more about the practices of the researcher, and their practice needs to live up to certain standards.

In the same way, it is entirely possible that experts perform *some* practices that allow them to counter the vitiating circumstances that naive subjects are affected by. But for that to be credible, we need specific proposals for vitiating circumstances that the experts are thought to be immune to (or, at least, better guarded against than lay subjects are), and we need information about the practices or other characteristics of the experts that make them immune.

This account casts light on one potential vitiating circumstance in particular, the mere exposure effect, although there are likely to be a number of others. To the best of my knowledge, there is no reason to expect experts to be immune to the mere exposure effect because of certain practices that they perform qua their expertise. On the contrary, naive subjects who are not immersed in language data from foreign languages and dialects, or who have not worked intensively on the specific construction in question will, though no less susceptible to this effect, have a smaller risk of being frequently exposed to relevant cases. And even if proponents of the expertise defence can make the case that experts perform certain practices that make them less susceptible to this effect than naive subjects, we need the reporting of results to be more transparent about these practices so that readers can evaluate whether the relevant precautions were taken or not.

To conclude, intuitive judgements reflect speakers' competence (mediated by their experience) when all goes well, but as subjects are not able to tell whether all is well or not in a given case, intuitive judgements cannot be used naively as direct expressions of speakers' competence. I do think, however, that it is questionable whether this is the current practice anyway. An interesting avenue for further studies would be a sociological exploration of what roles intuitive judgements in fact play in linguistic writing and argumentation and what precautions are actually taken when they are used as evidence for grammatical theories.

9.4 Perspectives

Let us turn now to one of the wider issues that form part of the backdrop for this work: the debate about the use of intuitive judgements as evidence in philosophy.

As for the Experimental Syntax debate, there is not unequivocal support for the use of experimental methods over the traditional method among generative linguists. This means that supporters of experimental methods in philosophy, like Machery and Stich (2013), who are enthusiastic about the rise of experimental methods in generative linguistics, may have to look elsewhere for support for using experimental methods in philosophy.[1] One option would be to look to non-generative linguistics, where there is general support for the use of experimental methods over the traditional armchair method. However, the results of the questionnaire study show that, unlike for the generative linguists, among the non-generative linguists there is not a majority of participants who think that intuitive judgements can stand alone as evidence in linguistics. So perhaps leaning on the non-generative linguists will not provide the required support for a defence of using experimental methods to collect intuitive judgements in philosophy either.

Whether the justification story that I present is applicable to intuitive judgements in philosophy is too big a question to take on in earnest here, but I will offer some initial speculation. On my account, linguistic intuitive judgements are ordinary judgements with no special input from the speaker's linguistic competence, but the speaker's linguistic competence does have a role to play in shaping the experience on which intuitive judgements are based. For the account to be applicable to intuitive judgements in philosophy, something similar would have to be the case there: The subject's "competence" with the topic in question would have to shape their experience of understanding the hypothetical scenario, and the judgement itself should then be based on the subject's accumulated experience of exercising their competence on that topic. On the face of it, such an account might work as a description of the intuitive judgements that are used as evidence in philosophy, though more work is needed to iron out the details. However, I am wondering whether it would also work to *justify* the use of intuitive judgements as evidence in philosophy. Although the process of arriving at an intuitive judgement might be the same in the case of linguistic intuitive judgements and philosophical intuitive judgements, the justification relies on the subject's "competence" with the topic in question being reliable.

It might be the case that we are as competent with *some* topics of interest

[1] See Schindler and Brøcker (forthcoming) for further discussion of the parallels between the experimental movements in philosophy and generative linguistics.

to philosophy as we are with our native languages,[2] but it is not clear to me that that is necessarily the case across the field. That would be a topic for further discussion and, ultimately, empirical investigation. See Nado (2014b), who argues that the mental states usually referred to in the philosophical literature as "intuitions" might in reality cover a broad range of states generated by heterogeneous mental processes. If that is the case, some of those processes are likely to be more reliable than others, meaning that the justification for using one kind as evidence does not extend to other kinds that are served by less reliable processes. See also Machery (2017, ch. 3), who argues that it is a non-accidental feature of the cases used to elicit intuitive judgements in philosophy that they pull apart what goes together under normal circumstances. This means that subjects cannot rely on their memory of similar cases to make intuitive judgements, which might undermine the applicability of my account to these cases and the justification it could potentially provide. This feature of the intuitive judgements used as evidence in philosophy would need to be discussed if one wanted to extend my account to justify this practice in philosophy.

Within philosophy, there is some discussion of what intuitive judgements are being used as evidence for. On the one side, some argue that they are used purely for conceptual analysis (figuring out what the content of the concepts that subjects use is; Machery 2017 argues that this is the way intuitive judgements should be used in philosophy). On the other side, some argue that intuitive judgements are, at least sometimes, really used to investigate more objective phenomena (Sosa, 2007; Bealer, 1998; Williamson, 2007). This discussion is contentious, and I do not want to go further into it here. I will note, however, that if one's aim is purely conceptual analysis, then the account I suggest would seem to offer some justification for the use of intuitive judgements as evidence, as we are surely competent with our own concepts. In that case, the story would be that subjects' "competence" with the concepts in question is what creates an immediate experience of whether that concept applies or not, and the judgement is then based on the accumulated previous experience subjects have with categorisation judgements of that kind.

Interestingly, intuitive judgements are used as evidence for theories in at least one other field as well, specifically in the context of thought experiments in physics. Can the use of intuitive judgements in linguistics offer any parallels for that field? Briefly put, I think similar reservations are relevant for the intuitive judgements that are used as evidence in thought experiments in physics with the exception that in physics, in general we are not interested in subjects' concepts

[2]Whether our linguistic competence is domain-specific and innately shaped or whether it is domain-general and input-based.

but rather in objective phenomena.[3]

Finally, let us return to the schism between generative and non-generative linguistics. In developing a justification for the use of intuitive judgements as evidence in linguistics, I have attempted to not rely on any assumptions about the nature of language that are specific to either of the main theoretical orientations in current linguistics. My account works whether our main interest is to study the abstract patterns observable in linguistic behaviour (although intuitive judgements in general are a less obvious choice for this than usage data is) or the underlying mental capacities. It does not take a stance on how language is acquired, whether we are born with language-specific innate rules, principles, or parameters, or whether we acquire language using general-domain learning mechanisms aided by our social abilities such as attention-reading. Since the account does not require subjects to apply their concept of grammaticality or the like, it is compatible with both the possibility that syntax is autonomous and with the alternative possibility that semantics and pragmatics are tightly interrelated with syntax. After eliciting the intuitive judgement, it is up to the researcher to interpret whether it is best explained by (purely) syntactic factors or whether both syntactic and semantic/pragmatic factors play a role and then further what to do with the conclusion. As mentioned in a previous chapter, I also do not intend to commit to a modular cognitive architecture, although I have used that vocabulary for ease of exposition and comparability with other accounts in the debate.

The empirical study sheds light on some of the different assumption underlying the practices of generative and non-generative linguists respectively. These results show both cases where the majority view differ among the two orientations and cases where the majority views are the same. There are also cases where generative linguists are significantly more likely to hold a particular view than non-generative linguists (and vice versa), regardless whether this view is the majority view among the generative linguists or not. This exposition of the theoretical beliefs of linguists from the different frameworks might, I hope, contribute in a small way to a better understanding of the motivations and assumptions of both positions across the divide. This approach of explicating the often undiscussed theoretical assumptions underlying research practices could be adapted and applied for other issues where the two sides are in disagreement. While it takes effort to get to understand the other side, especially when one is not familiar with technical terminology, historical references, or indeed the motivating assumptions of the other side, in the long run linguistics as a field would benefit from a greater exchange of ideas across theoretical boundaries.

[3]See Schindler and Saint-Germier (under review) for a comparison between intuitive judgements elicited by thought experiments in physics and philosophy respectively.

9.5 Summing up

This dissertation provides an answer to the open descriptive question that has been one of two central issues in the etiology debate: Do a majority of generative linguists subscribe to VoC? As we have seen, the answer turned out to be negative, but the majority generative view is not one of the other main views defended in that debate either. With this question out of the way, more attention can be directed to the normative question at the centre of the etiology debate: What position should we take on the justification of the use of intuitive judgements as evidence for grammatical theories? My contribution to this part of the debate is an account that, I argue, neither reduces to VoC nor to the Modest Explanation. Instead, it places speakers' experience of hearing the sentence at the centre, giving both the speakers' competence and their central processor important roles to play.

Part IV

Bibliography and appendices

Bibliography

Adger, D. (2015). Syntax. *WIREs Cognnitive Science*, 6:131–147.

Ambridge, B. and Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics*, 19(3):357–389.

Armstrong, S. L., Gleitman, L. and Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13:263–308.

Bard, E. G., Robertson, D. and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*:32–68.

Bauer, M. W. (2000). Classical content analysis: A review. In Bauer, M. W. and Gaskell, G. (eds.), *Qualitative researching with text, image and sound: A practical handbook*. London: SAGE Publications, 131–151.

Bealer, G. (1998). Intuition and the autonomy of philosophy. In DePaul, M. and Ramsey, W. (eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry*. Lanham, MD: Rowman & Littlefield, 201–240.

Bermúdez, J. L. (2010). *Cognitive science: An introduction to the science of the mind*. Cambridge: Cambridge University Press.

Botha, R. P. (1981). *The conduct of linguistic inquiry*. New York: Mouton.

Botha, R. P. (1982). On 'the Galilean style' of linguistic inquiry. *Lingua*, 58(1):1–50.

Brøcker, K. (forthcoming). Do generative linguists believe in a Voice of Competence? In Schindler, S., Drożdżowicz, A. and Brøcker, K. (eds.), *Linguistic intuitions*. Oxford: Oxford University Press.

Butler, C. S. (2003). *Structure and function: A guide to three major structural-functional theories: Part 1: Approaches to the simplex clause*, volume 63 of *Studies in Language Companion Series*. Amsterdam: John Benjamins Publishing.

Butler, C. S. (2006). Functionalist theories of language. In Brown, K. (ed.), *Encyclopedia of Language & Linguistics*. Oxford: Elsevier, 2nd edition, 696–704.

Carr, P. (2003). Innateness, internalism and input: Chomskyan rationalism and its problems. *Language Sciences*, 25(6):615–635.

Chelliah, S. (2013). Fieldwork for language description. In Podesva, R. J. and Sharma, D. (eds.), *Research methods in linguistics*. Cambridge: Cambridge University Press, 51–73.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press, 1st edition.

Chomsky, N. (1973). Conditions on transformations. In Anderson, S. R. and Kiparsky, P. (eds.), *A festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 232–286.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, N. (1995). Language and nature. *Mind*, 104(413):1–61.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates, 2nd edition.

Collins, J. (2008). Knowledge of language redux. *Croatian Journal of Philosophy*, 8(1):3–34.

Cowart, W. (1997). *Experimental syntax – Applying objective methods to sentence judgments*. Thousand Oaks, CA: SAGE Publications, 1st edition.

Culbertson, J. and Gross, S. (2009). Are linguists better subjects? *The British Journal for the Philosophy of Science*, 60(4):721–736.

Culicover, P. W. and Jackendoff, R. (2010). Quantitative methods alone are not enough: Response to Gibson and Fedorenko. *Trends in Cognitive Sciences*, 14(6):234–235.

Darnell, M., Moravcsik, E. A., Noonan, M., Newmeyer, F. J. and Wheatley, K. (1999a). *Functionalism and formalism in linguistics: Volume I: General papers*, volume 41 of *Studies in language companion series*. Amsterdam: John Benjamins Publishing Company.

Darnell, M., Moravcsik, E. A., Noonan, M., Newmeyer, F. J. and Wheatley, K. (1999b). *Functionalism and formalism in linguistics: Volume II: Case studies*, volume 42 of *Studies in language companion series*. Amsterdam: John Benjamins Publishing Company.

Dąbrowska, E. (2008). Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics*, 19(3):391–425.

Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *Linguistic Review*, 27(1):1–23.

Dąbrowska, E. (2013). Functional constraints, usage, and mental grammars: A study of speakers' intuitions about questions with long-distance dependencies. *Cognitive Linguistics*, 24(4):633–665.

Dąbrowska, E. (2016). Looking into introspection. In Drożdż, G. (ed.), *Studies in Lexicogrammar: Theory and applications*. Amsterdam: John Benjamins Publishing Company, 55–74.

Den Dikken, M., Bernstein, J. B., Tortora, C. and Zanuttini, R. (2007). Data and grammar: Means and individuals. *Theoretical Linguistics*, 33(3):335–352.

Devitt, M. (2006a). Defending Ignorance of language: Responses to the Dubrovnik Papers. *Croatian Journal of Philosophy*, VI(18):571–606.

Devitt, M. (2006b). *Ignorance of language*. Oxford: Oxford University Press.

Devitt, M. (2006c). Intuitions in linguistics. *The British Journal for the Philosophy of Science*, 57(3):481–513.

Devitt, M. (2010a). Linguistic intuitions revisited. *The British Journal for the Philosophy of Science*, 61(4):833–865.

Devitt, M. (2010b). What "intuitions" are linguistic evidence? *Erkenntnis*, 73(2):251–264.

Devitt, M. (2011). Experimental semantics. *Philosophy and Phenomenological Research*, 82(2):418 – 435.

Devitt, M. (2012). Whither experimental semantics? *Theoria*, 27(1):5–36.

Devitt, M. (2013). Linguistic intuitions are not "the Voice of Competence". In Haug, M. C. (ed.), *Philosophical methodology: The armchair or the laboratory?* New York: Routledge, 268–293.

Devitt, M. (2015). Testing theories of reference. In Haukioja, J. (ed.), *Advances in experimental philosophy of language*. London: Bloomsbury Academic, 31–63.

Diamantopoulos, A. and Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2):269–277.

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Drożdżowicz, A. (2018). Speakers' intuitive judgements about meaning – the Voice of Performance view. *Review of Philosophy and Psychology*, 9(1):177–195.

Fanselow, G. (2007). Carrots – perfect as vegetables, but please not as a main dish. *Theoretical Linguistics*, 33(3):353–367.

Featherston, S. (2007). Data in generative grammar: The stick and the carrot. *Theoretical Linguistics*, 33(3):269–318.

Featherston, S. (2009). Relax, lean back, and be a linguist. *Zeitschrift für Sprachwissenschaft*, 28(1):127–132.

Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review*, 22(2-4):365–380.

Ferreira, F., Christianson, K. and Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of psycholinguistic research*, 30(1):3–20.

Fiengo, R. (2003). Linguistic intuitions. *The Philosophical Forum*, 34(3-4):253–266.

Fischer, E. and Engelhardt, P. E. (2016). Intuitions' linguistic sources: Stereotypes, intuitions and illusions. *Mind & Language*, 31(1):67–103.

Fitzgerald, G. (2010). Linguistic intuitions. *The British Journal for the Philosophy of Science*, 61(1):123–160.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: The MIT Press.

Fox, B. A., Hayashi, M. and Jasperson, R. (1996). Resources and repair: A cross-linguistic study of syntax and repair. In Ochs, E., Schegloff, E. A., Thompson, S. A. and Thompson, S. A. (eds.), *Interaction and grammar*, volume 13 of *Studies in interactional sociolinguistics*. Cambridge: Cambridge University Press, 185–237.

Gibbs Jr., R. W. (2006). Introspection and cognitive linguistics: Should we trust our own intuitions? *Annual Review of Cognitive Linguistics*, 4(1):135–151.

Gibson, E. and Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, 14(6):233–234.

Gibson, E. and Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2):88–124.

Gibson, E., Piantadosi, S. and Fedorenko, K. (2011). Using Mechanical Turk to obtain and analyze English acceptability judgments: Linguistic acceptability on Mechanical Turk. *Language and Linguistics Compass*, 5(8):509–524.

Gibson, E., Piantadosi, S. T. and Fedorenko, E. (2013). Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes*, 28(3):229–240.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gordon, P. C. and Holyoak, K. J. (1983). Implicit learning and generalization of the "mere exposure" effect. *Journal of Personality and Social Psychology*, 45(3):492–500.

Greenbaum, S. (1976). Contextual influence on acceptability judgements. *Linguistics*, 15(187):5–11.

Grewendorf, G. (2007). Empirical evidence and theoretical reasoning in generative grammar. *Theoretical Linguistics*, 33(3):369–380.

Gries, S. T. (2014). Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. In Gisborne, N. and Hollmann, W. B. (eds.), *Theory and data in cognitive linguistics*. Amsterdam: John Benjamins Publishing Company, 15–48.

Gross, S. (forthcoming). Linguistic intuitions: Error signals and the Voice of Competence. In Schindler, S., Drożdżowicz, A. and Brøcker, K. (eds.), *Linguistic intuitions*. Oxford: Oxford University Press.

Gross, S. and Culbertson, J. (2011). Revisited linguistic intuitions. *The British Journal for the Philosophy of Science*, 62(3):639–656.

Haider, H. (2007). As a matter of facts – comments on Featherston's sticks and carrots. *Theoretical Linguistics*, 33(3):381–394.

Haider, H. (2009). The thin line between facts and fiction. In Winkler, S. and Featherston, S. (eds.), *The fruits of emperical linguistics: Volume 1: Process*, volume 101 of *Studies in generative grammar*. New York: Mouton de Gruyter, 75–102.

Hansen, E. and Heltoft, L. (2011). *Grammatik over det danske sprog*. København: Det Danske Sprog- og Litteraturselskab.

Hansen, K. M. (2010). Spørgeskemadesign. In Andersen, L. B., Hansen, K. M. and Klemmensen, R. (eds.), *Metoder i statskundskab*. København: Hans Reitzels Forlag, 1st edition, 282–303.

Hellevik, O. (2002). *Forskningsmetode i sosiologi og statsvitenskap*. Oslo: Universitetsforlaget AS, 7th edition.

Hindle, D. and Sag, I. (1975). Some more on anymore. In Fasold, R. W. and Shuy, R. W. (eds.), *Analyzing variation in language: Papers from the second colloquium on new ways of analyzing variation*. Washington, DC: Georgetown University Press, 89–110.

Karanth, P., Kudva, A. and Vijayan, A. (1995). Literacy and linguistic awareness. In de Gelder, B. and Morais, J. (eds.), *Speech and reading: A comparative approach*. Hove: Erlbaum (UK) Taylor and Francis, 303–316.

Kayne, R. S. (1983). Connectedness. *Linguistic inquiry*, 14(2):223–249.

Kieras, D. (2007). Control of cognition. In Gray, W. D. (ed.), *Integrated models of cognitive systems*. Oxford: Oxford University Press, 327–355.

Kinzler, K. D., Shutts, K. and Spelke, E. S. (2012). Language-based social preferences among children in South Africa. *Language learning and development*, 8(3):215–232.

Labov, W. (1973). Where do grammars stop? In *23rd annual round table*, volume 25 of *Monograph series on language and linguistics*. 43–88.

Labov, W. (1975). Empirical foundations of linguistic theory. In Austerlitz, R. (ed.), *The scope of American linguistics*, The first golden anniversary symposium of the Linguistic Society of America. Ghent: The Peter de Ridder Press, 77–133.

Lewis, D. (1983). *Philosophical papers: Volume I.* New York: Oxford University Press.

Ludlow, P. (2011). *The philosophy of generative linguistics.* Oxford: Oxford University Press.

Luka, B. (2005). A cognitively plausible model of linguistic intuitions. In *Polymorphous linguistics: Jim McCawley's legacy.* Cambridge, MA: The MIT Press, 479–502.

Luka, B. and Barsalou, L. (2005). Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language*, 52(3):436–459.

Machery, E. (2011). Thought experiments and philosophical knowledge. *Metaphilosophy*, 42(3):191–214.

Machery, E. (2017). *Philosophy within its proper bounds.* Oxford: Oxford University Press.

Machery, E., Mallon, R., Nichols, S. and Stich, S. P. (2013). If folk intuitions vary, then what? *Philosophy and Phenomenological Research*, 86(3):618–635.

Machery, E. and Stich, S. (2013). The role of experiment in the philosophy of language. In Russell, G. and Fara, D. G. (eds.), *Routledge companion to philosophy of language.* New York: Routledge, 495–512.

Majid, A. (2012). A guide to stimulus-based elicitation for semantic categories. In Thieberger, N. (ed.), *The Oxford handbook of linguistic fieldwork.* Oxford: Oxford University Press, 54–71.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Fransisco: W.H. Freeman and Company.

Marvasti, A. (2004). *Qualitative research in sociology.* Thousand Oaks, CA: SAGE Publications.

Maynes, J. (2012a). Linguistic intuition and calibration. *Linguistics and Philosophy*, 35(5):443–460.

Maynes, J. (2012b). *Meaningful intuitions: the evidential role of intuitions in the study of language.* PhD thesis. Baltimore, MD: Johns Hopkins University.

Maynes, J. and Gross, S. (2013). Linguistic intuitions. *Philosophy Compass*, 8(8):714–730.

Miščević, N. (2006). Intuitions: The discrete Voice of Competence. *Croatian Journal of Philosophy*, 6(18):523–548.

Morgan, D. L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research*, 3(1):112–121.

Morgan, D. L. (2014). *Integrating qualitative & quantitative methods – A pragmatic approach.* Thousand Oaks, CA: SAGE Publications.

Myers, J. (2009a). Automated collection and analysis of phonological data. In Winkler, S. and Featherston, S. (eds.), *The fruits of empirical linguistics: Volume 1: Process*, volume 101 of *Studies in generative grammar*. Berlin: De Gruyter Mouton, 151–176.

Myers, J. (2009b). Syntactic judgment experiments. *Linguistics and Language Compass*, 3(1):406–423.

Nado, J. (2014a). Philosophical expertise. *Philosophy Compass*, 9(9):631–641.

Nado, J. (2014b). Why intuition? *Philosophy and Phenomenological Research*, 89(1):15–41.

Newmeyer, F. J. (1983). *Grammatical theory: Its limits and possibilities*. Chicago: University of Chicago Press.

Newmeyer, F. J. (1991). Functional explanation in linguistics and the origins of language. *Language & Communication*, 11(1-2):3–28.

Newmeyer, F. J. (1998). *Language form and language function*. Cambridge, MA: The MIT Press.

Newmeyer, F. J. (2007). Commentary on Sam Featherston, 'Data in generative grammar: The stick and the carrot'. *Theoretical Linguistics*, 33(3):395–399.

Nicholls, J. G., Licht, B. G. and Pearl, R. A. (1982). Some dangers of using personality questionnaires to study personality. *Psychological Bulletin*, 92(3):572–580.

Olsen, H. (2006). *Guide til gode spørgeskemaer*. København: Socialforskningsinstituttet.

Petersen, M. B. (2010). Indekskonstruktion. In Andersen, L. B., Hansen, K. M. and Klemmensen, R. (eds.), *Metoder i statskundskab*. København: Hans Reitzels Forlag, 1st edition, 393–413.

Phillips, C. (2009). Should we impeach armchair linguists? *Japanese/Korean Linguistics*, 17:1–16.

Phillips, C. and Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. In Gaskell, M. G. (ed.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, 739–756.

Pietroski, P. (2008). Think of the children. *Australasian Journal of Philosophy*, 86(4):657–669.

Pust, J. (2017). Intuition. In Zalta, E. N. (ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition. Accessed: January 30, 2019.
URL `https://plato.stanford.edu/archives/sum2017/entries/intuition/`

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1):20–43.

Rey, G. (2013). The possibility of a naturalistic Cartesianism regarding intuitions and introspection. In Haug, M. C. (ed.), *Philosophical methodology: The armchair or the laboratory ?* New York: Routledge, 243–267.

Rey, G. (forthcoming). *Representation of language: Foundational issues in a Chomskyan linguistics.* Oxford: Oxford University Press. Accessed: January 30, 2019.
URL `https://sites.google.com/site/georgesrey/`

Rosengren, K. E. (1981). Advances in Scandinavian content analysis: An introduction. In Rosengren, K. E. (ed.), *Advances in content analysis*, volume 9 of *SAGE Annual reviews of comminication research.* London: SAGE Publications, 9–19.

Sampson, G. (2001). *Empirical linguistics.* London: Continuum.

Santana, C. (forthcoming). A non-evidential case for the use of intuitions in linguistics. In Schindler, S., Drożdżowicz, A. and Brøcker, K. (eds.), *Linguistic Intuitions.* Oxford: Oxford University Press.

Schegloff, E. A., Jefferson, G. and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Schindler, S. and Brøcker, K. (forthcoming). Experiments in syntax and philosophy: The method of choice? In Schindler, S., Drożdżowicz, A. and Brøcker, K. (eds.), *Linguistic intuitions.* Oxford: Oxford University Press.

Schindler, S. and Saint-Germier, P. (under review). Armchair physics and the method of cases.

Schütze, C. T. (2006/1996). *The empirical base of linguistics: Grammaticality judgements and linguistic methodology.* Classics in linguistics 2. Berlin: Language Science Press. First published 1996 by The University of Chicago Press.

Schütze, C. T. (2006). Data and evidence. In Brown, K. (ed.), *Encyclopedia of Language & Linguistics*, volume 3. Oxford: Elsevier, 2nd edition, 356–363.

Schütze, C. T. (2011). Linguistic evidence and grammatical theory. *WIREs Cognitive Science*, 2:206–221.

Schütze, C. T. and Sprouse, J. (2013). Judgment data. In Podesva, R. J. and Sharma, D. (eds.), *Research methods in linguistics.* Cambridge: Cambridge University Press, 27–50.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual review of psychology*, 46(1):561–584.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures.* Boca Raton: crc Press, 3rd edition.

Smith, N. (1999). *Chomsky: Ideas and ideals.* Cambridge: Cambridge University Press.

Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, 132(1):99–107.

Spector, P. E. (1992). *Summated rating scale construction: An introduction.* Newbury Park, CA: SAGE Publications.

Sprouse, J. (2013). Acceptability judgments. *Oxford Bibliographies Online: Linguistics*. Accessed: January 30, 2019.
URL `http://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0097.xml`

Sprouse, J. (2015). Three open questions in experimental syntax. *Linguistics Vanguard*, 1(1):89–100.

Sprouse, J. (2018). Acceptability judgments and grammaticality, prospects and challenges. In Hornstein, N., Lasnik, H., Patel-Grosz, P. and Yang, C. (eds.), *Syntactic Structures after 60 years: The impact of the Chomskyan revolution in linguistics*, volume 129 of *Studies in generative grammar*. Berlin: De Gruyter Mouton, 195–223.

Sprouse, J. and Almeida, D. (2012a). Assessing the reliability of textbook data in syntax: Adger's Core syntax. *Journal of Linguistics*, 48(03):609–652.

Sprouse, J. and Almeida, D. (2012b). Power in acceptability judgment experiments and the reliability of data in syntax. Unpublished manuscript. Accessed: January 30, 2019.
URL `http://ling.auf.net/lingbuzz/001520`

Sprouse, J. and Almeida, D. (2013a). The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*, 28(3):222–228.

Sprouse, J. and Almeida, D. (2013b). The role of experimental syntax in an integrated cognitive science of language. In Boeckx, C. and Grohmann, K. K. (eds.), *The Cambridge handbook of biolinguistics*. Cambridge: Cambridge University Press, 181–202.

Sprouse, J., Schütze, C. T. and Almeida, D. (2011). Assessing the reliability of journal data in syntax: Linguistic Inquiry 2001-2010. Unpublished manuscript. Accessed: January 30, 2019.
URL `http://ling.auf.net/lingbuzz/001352`

Sprouse, J., Schütze, C. T. and Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua*, 134:219–248.

Textor, M. (2009). Devitt on the epistemic authority of linguistic intuitions. *Erkenntnis*, 71(3):395–405.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

van Inwagen, P. (1997). Materialism and the psychological-continuity account of personal identity. *Philosophical Perspectives*, 11:305–319.

Wasow, T. and Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, 115(11):1481–1496.

Weinberg, J. M., Gonnerman, C., Buckner, C. and Alexander, J. (2010). Are philosophers expert intuiters? *Philosophical Psychology*, 23(3):331–355.

Willems, K. (2012). Intuition, introspection and observation in linguistic inquiry. *Language sciences*, 34(6):665–681.

Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science*, 10(6):224–228.

Appendix A

Questionnaire

The questionnaire described in part II of the dissertation is presented in full, including instructions, on the following pages. First, the call for participants (posted on LinguistList) is presented.

If a participant chose the option "...can in no way be used as evidence" in the first question, they were redirected to the end of the questionnaire, where they were given the option to sign up for updates and for the lottery (see section 5.5). The question labelled "The form of implementation of mental rules" was only displayed to participants who chose either "It is a good hypothesis that structure rules are actually implemented in the minds of speakers" or "[Supply other answer]" in the previous question ("Structure rules and mental rules") as described in section 6.1.2.7. A few of the background questions were also only displayed depending on the participants' previous answers to avoid redundant or irrelevant questions.

APPENDIX A: QUESTIONNAIRE

LINGUIST List 28.2931

Tue Jul 04 2017

Qs: Survey: intuitions as evidence in linguistics

Editor for this issue: Sarah Robinson <srobinson@linguistlist.org&gt

---

Date: 04-Jul-2017
From: Karen Brøcker <karenbroecker@css.au.dk>
Subject: Survey: intuitions as evidence in linguistics
E-mail this message to a friend

Dear colleagues,

In the past decades, there has been lively debates over the way native speakers' intuitive judgements are used as evidence in linguistics. With this questionnaire, we want to explore the experiences and opinions of members of the research community about this issue.

More specifically, we are interested in the opinions of researchers ...
- who believe that native speakers' intuitive judgements of syntactic well-formedness can, to some extent, be used as evidence in linguistics (e.g. for fieldwork, grammatical descriptions, theoretical work etc.),
- who subscribe to any theoretical background (formal/generative, functional/cognitive, other), and
- who have obtained or are studying for a PhD in linguistics or closely related fields.

Your help would be highly appreciated! The questionnaire takes approximately 25 minutes to complete. Upon completing it, participants have the option to enter in a lottery for ten vouchers for amazon.com, each worth $25.

Please follow this link to participate: https://aarhus.eu.qualtrics.com/jfe/form/SV_39FUiioVRCK8JtX

You are also more than welcome to forward this call to any colleague(s) who meet the requirements and who might be interested in participating in the survey.

This survey is part of the project Intuitions in Science and Philosophy: www.projects.au.dk/intuitions/. You can contact karenbroecker@css.au.dk for more information about the survey.

Kind regards,
Karen Brøcker, PhD fellow
Aarhus University, Centre for Science Studies

Linguistic Field(s): Cognitive Science
Discipline of Linguistics
General Linguistics
Language Documentation
Linguistic Theories
Philosophy of Language
Syntax

Page Updated: 04-Jul-2017

Instructions

Syntactic intuitions as evidence in linguistics

In this survey, we would like to ask about your experiences and opinions regarding a type of linguistic data that is sometimes called "syntactic intuitions", "acceptability judgements", or "grammaticality judgements". These are taken to say something about the *morphosyntactic well-formedness of sentences*. For example, most native speakers of English would likely agree that "the cat is on the table" is a well-formed sentence of English, whereas "cat table the on is" is not. We are interested in what you think about syntactic intuitions as *evidence* for descriptions or theories of grammar. Your responses will be anonymous.

You will be presented with a number of statements. In some cases, you will be asked to indicate to which extent you agree with a statement. In other cases, you will be asked to choose from a list the option that best expresses your opinion. Choose "[Supply other answer]" and provide a clarification if none of the options express your opinion. Use as much time as you need.

Syntactic intuitions and other types of linguistic evidence

Syntactic intuitions and other types of linguistic evidence

Please choose the option that best expresses your opinion:

When studying grammatical phenomena, syntactic intuitions ...

- ○ … can in some cases stand alone as evidence.
- ○ … can in some cases serve as evidence but can never stand alone.
- ○ … can in no way be used as evidence.
- ○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

The subject matter of the study of grammar

The subject matter of the study of grammar

Please choose the option that best expresses your opinion:

When I study grammatical phenomena, I ultimately seek to understand...

- ○ … the systematic patterns found in linguistic behaviour.
- ○ … the linguistic capacity of the mind.
- ○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Acceptability and grammaticality

Acceptability and grammaticality

Most native speakers of English would find the sentence "the dog that the woman that the man saw owned ran" to be an intuitively unnatural sentence in English. In the terminology of generative linguistics, the sentence is *unacceptable* to those speakers (this is not to be confused with whether or not the sentence is infelicitous to the speakers in some specific context). Most native speakers would also most likely find the sentence intuitively *ungrammatical*. However, some experts argue that it is, in fact, grammatical.

Please choose the option that best expresses your opinion:

To the extent syntactic intuitions can serve as evidence for theories of grammar, only those syntactic intuitions can serve as evidence that are …

- ○ … acceptability intuitions.
- ○ … grammaticality intuitions.
- ○ [Supply other answer] _____

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

_____

The origin of syntactic intuitions

The origin of syntactic intuitions

Please choose the option that best expresses your opinion:

When syntactic intuitions are reliable as evidence, this is *mainly* because ...

- ○ … they are speakers' *reflections* about language use, and speakers are to some degree reliable judges about this.
- ○ … they express speaker's *competence* in their native language.

○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Constraints and mental rules

Structure rules and mental rules

Syntactic intuitions is one type of evidence used by linguists to characterise how languages are structured. Let's call the rules that describe this the *structure rules* of particular languages.

Please choose the option that best expresses your opinion:

The structure rules that linguists describe are sometimes said to be "implemented in the minds of speakers".

- ○ It is a good hypothesis that structure rules are actually implemented in the minds of speakers.
- ○ From the structure rules we observe, we can only infer that the mind works *as if* it was following those rules.
- ○ From the structure rules we observe, we cannot infer anything about how the mind processes language.
- ○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

The form of implementation of mental rules

Please choose the option that best expresses your opinion:

There must be something in the mind that gives rise to what we call "rules of grammar".

- ○ The rules of grammar are probably explicitly represented in the mind. If one could look into subjects' minds, one could find explicit rules.
- ○ The rules of grammar are implemented in the mind, but they are probably not explicitly represented.
- ○ The rules of grammar are probably not implemented in the mind.
- ○ [Supply other answer] _____

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

_____

The role of mental grammar

The role of mental grammar

Please choose the option that best expresses your opinion:

Syntactic intuitions are sometimes said to be "deduced from the speaker's mental grammar".

- ○ This is probably a poor description of how intuitions are formed.
- ○ This is a good way to talk about how intuitions are formed but should probably not be taken too literally.
- ○ This is likely to be the actual process of how syntactic intuitions are formed in the mind.
- ○ [Supply other answer] _____

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

_____

Fallibility of intuitions

Fallibility of intuitions

To what extent do you agree with the **second** statement below:

Imagine you could abstract away all performance factors that might influence a speaker's syntactic intuitions.

In that case, it would be possible for the resulting syntactic intuitions to be mistaken about the grammatical properties of the sentence.

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
○	○	○	○	○

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Experts and ordinary speakers

Experts and ordinary speakers

Assume here that both linguists and ordinary speakers are native speakers of the same language.

Please choose the option that best expresses your opinion:

The syntactic intuitions of linguists working on theories of grammar are, *on average*, …

- ○ … worse evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
- ○ … equally good evidence for theories of grammar as the syntactic intuitions of ordinary speakers.
- ○ … better evidence for theories of grammar than the syntactic intuitions of ordinary speakers.
- ○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Methodological approach

Methodological approach

To what extent do you agree with the following statement:

In general, consulting one's own or one's colleague's syntactic intuitions produces good evidence for theories of grammar.

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
○	○	○	○	○

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Methodological approach

To what extent do you agree with the following statement:

In general, syntactic intuitions should be collected and analysed by experimental methods from large numbers of speakers and using statistical tests.

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
○	○	○	○	○

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Methodological approach

To what extent do you agree with the following statement:

In general, if one has to choose, it is more important that a theory is built on reliable data than that it lives up to theoretical virtues such as simplicity, elegance, and fruitfulness.

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
○	○	○	○	○

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Methodological approach

Please choose the option that best expresses your opinion:

Some linguists use gradient rather than binary scales to collect syntactic intuitions.

- ○ Well-designed gradient scales may very well reflect real degrees of grammaticality.
- ○ Even well-designed gradient scales probably just capture effects that are not due to grammaticality.
- ○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Significance

Significance

Please choose the answer that best expresses your situation:

Can you think of a situation, before answering this survey, where you thought about whether syntactic intuitions can serve as evidence for theories of grammar?

- ○ Yes
- ○ No
- ○ [Supply other answer]

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Significance

To what extent do you agree with the following statement:

Linguists who use intuitions as evidence should set aside time to consider *why* intuitions can serve as evidence for their theories.

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
○	○	○	○	○

Optional: Would you like to elaborate on your answer or comment on anything else regarding this section?

Powered by Qualtrics

Background questions

We will now ask you some questions about your background.

Please answer all of them. *Without your answers the questionnaire will be incomplete.*

What is your current position? [choose 1]

- ○ PhD student
- ○ Postdoc
- ○ Non-tenure-track academic position
- ○ Tenure-track academic position
- ○ Other [please specify]: _____

Do you hold a doctorate?

- ○ Yes
- ○ No

Is your research mainly in linguistics?

- ○ Yes
- ○ No. My research is mainly in [please specify]: _____
- ○ I don't do research.

What are your main areas of research within linguistics? [choose 1-3]

- ☐ Morphology
- ☐ Syntax
- ☐ Phonetics
- ☐ Phonology
- ☐ Semantics
- ☐ Pragmatics
- ☐ Sociolinguistics
- ☐ Anthropological linguistics
- ☐ Psycholinguistics

- [ ] Cognitive linguistics
- [ ] Typology
- [ ] Historical linguistics
- [ ] Theoretical linguistics
- [ ] Language acquisition
- [ ] Text linguistics
- [ ] Corpus linguistics
- [ ] Other [please specify]: _____

Please choose the option that best describes your theoretical perspective on language: [choose 1]

- ○ Formal/generative
- ○ Functional/cognitive
- ○ Other/atheoretical [please specify] _____
- ○ Mixed [please specify, which approach(es), and which one (if any) do you mostly lean towards?] _____

Name two or three well-known linguists that exemplify your theoretical perspective on language (e.g., Chomsky, Saussure, Halliday, etc.):

_____

Have you ever used syntactic intuitions as evidence in your own research?

- ○ Yes
- ○ No

Please check all that apply:

- [ ] I have used my own syntactic intuitions.
- [ ] I have used the syntactic intuitions of other informants than myself.

Please check all that apply:

- [ ] I have used syntactic intuitions collected informally (in the armchair).

- [ ] I have used syntactic intuitions collected through linguistic fieldwork.
- [ ] I have used syntactic intuitions collected through surveys.
- [ ] Other [please specify]: _____

In what region of the world do you currently live? (if you live in multiple places, where do you spend most time of the year?) [choose 1]

- ○ Africa
- ○ Asia
- ○ Europe
- ○ Central and South America incl. the Caribbean
- ○ Oceania, incl. Australasia
- ○ Canada and the USA
- ○ Other [please specify]: _____

Please choose the option that best describes you: [choose 1]

- ○ Female
- ○ Male
- ○ Other

How old are you? [choose 1]

- ○ 30 or under
- ○ 31-40
- ○ 41-50
- ○ 51-60
- ○ 61 or over

How would you rate your level of proficiency in English? [choose 1]

- ○ Beginner
- ○ Intermediate
- ○ Advanced
- ○ Native

Any other comments

Any other comments?

Optional: We welcome any comments you might have about specific questions or further issues relating to the topic of the questionnaire:

Lottery, future participation, and updates

Lottery, future participation, and updates

This is the last slide of the questionnaire. Below, you have the option to supply your e-mail address and indicate whether you would be interested in 1) entering our lottery, 2) participating in a future part of the study, and/or 3) receiving an update about the results of this questionnaire.

If you do not wish to share your e-mail address, that is completely fine. In case you do, your e-mail address will ONLY be used for the purpose(s) selected by you, it will not be shared with any third party, and it will be erased from the data set at the end of the study. Your responses will remain confidential.

- ○ I do NOT want to share my e-mail address for any of these purposes.
- ○ I agree to sharing my e-mail address, and my e-mail address is:

My e-mail address may be used for the following purpose(s): [check all that apply]

- ☐ To enter a lottery for ten vouchers for amazon.com, each worth $25.
- ☐ To send out an enquiry for participation in a round of interviews to be conducted as a future part of this study (winter 2017 or spring 2018).
- ☐ To send an update about the results of this survey.

Powered by Qualtrics

Appendix B

Chi-square tests

Overview tables for all the chi-square Goodness of Fit tests performed for this dissertation are presented on the following pages. The method of analysis is described in section 6.1.1.

An alpha-level of .05 was used for all statistical tests, excluding the pairwise comparisons (post-hoc tests), where the relevant Bonferroni corrections were applied. In the cases where three comparisons were performed the corrected alpha-level is $.05/3 = .017$, and in the cases where six comparisons were performed the corrected alpha-level is $.05/6 = .008$.

The questions are presented in the same order here as in section 6.2. First, the generative participants' results for questions relating to the etiology debate are presented, then their results for questions relating to the Experimental Syntax debate, and finally their results for other questions. For the generative participants, results are presented for both the large group including all generative participants and for the smaller group who identified exclusively as formal/generative (see section 5.5 for an introduction to the groups). The non-generative participants' results are then presented in the same order.

B.1 Chi-square tests: generative participants, the etiology debate

The origin of syntactic intuitions:

All generative participants:

Answer	Reflections	Competence	Other
Frequency	17 (23.29%)	40 (55.79%)	16 (22.92%)
Chi-square GOF	$\chi^2(2, n=73) = 15.15, p < .001, V = .32$		

Pairwise comparisons:

Comparison	Reflections vs Competence	Reflections vs Other	Competence vs Other
Chi-square GOF	$\chi^2(1, n=57) = 9.28$, $p = .002, V = .40$	$\chi^2(1, n=33) = .03$, $p = .862$ (ns)	$\chi^2(1, n=56) = 10.29$, $p = .001, V = .43$

Small generative group:

Answer	Reflections	Competence	Other
Frequency	12 (21.05%)	34 (59.65%)	11 (19.30%)
Chi-square GOF	$\chi^2(2, n=57) = 17.80, p < .001, V = .40$		

Pairwise comparisons:

Comparison	Reflections vs Competence	Reflections vs Other	Competence vs Other
Chi-square GOF	$\chi^2(1, n=46) = 10.52$, $p = .001, V = .48$	$\chi^2(1, n=23) = .04$, $p = .835$ (ns)	$\chi^2(1, n=45) = 11.76$, $p < .001, V = .51$

Acceptability and grammaticality:

All generative participants:

Answer	Acceptability	Grammaticality	Other
Frequency	31 (42.47%)	17 (23.29%)	25 (34.25%)
Chi-square GOF	$\chi^2(2, n=73) = 4.05, p = .132$ (ns)		

Small generative group:

Answer	Acceptability	Grammaticality	Other
Frequency	24 (42.11%)	14 (24.56%)	19 (33.33%)
Chi-square GOF	$\chi^2(2, n=57) = 2.63, p = .268$ (ns)		

The role of mental grammar:

All generative participants:

Answer	Poor description	Not literal	Other	Actual process
Frequency	13 (17.81%)	44 (60.27%)	8 (10.96%)	8 (10.96%)
Chi-square GOF	$\chi^2(3, n=73) = 49.36, p < .001, V = .47$			

Pairwise comparisons:

Comparison	Poor description vs Not literal	Poor description vs Other	Poor description vs Actual process
Chi-square GOF	$\chi^2(1, n=57) = 16.86$, $p < .001, V = .54$	$\chi^2(1, n=21) = 1.19$, $p = .275$ (ns)	$\chi^2(1, n=21) = 1.19$, $p = 0.275$ (ns)

Comparison	Not literal vs Other	Not literal vs Actual process	Other vs Actual process
Chi-square GOF	$\chi^2(1, n=52) = 24.92$, $p < .001, V = .69$	$\chi^2(1, n=52) = 24.92$, $p < .001, V= .69$	Not compared

Small generative group:

Answer	Poor description	Not literal	Other	Actual process
Frequency	10 (17.54%)	33 (57.89%)	7 (12.28%)	7 (12.28%)
Chi-square GOF	$\chi^2(3, n=57) = 33.32, p < .001, V = .44$			

Pairwise comparisons:

Comparison	Poor description vs Not literal	Poor description vs Other	Poor description vs Actual process
Chi-square GOF	$\chi^2(1, n=43) = 12.30$, $p < .001, V = .53$	$\chi^2(1, n=17) = 0.53$, $p = .467$ (ns)	$\chi^2(1, n=17) = 0.53$, $p = .467$ (ns)

Comparison	Not literal vs Other	Not literal vs Actual process	Other vs Actual process
Chi-square GOF	$\chi^2(1, n=40) = 16.9$, $p < .001, V = .65$	$\chi^2(1, n=40) = 16.9$, $p < .001, V = .65$	Not compared

Fallibility of intuitions:[1]

All generative participants:

Answer	Agree	Neither	Disagree
Frequency	18 (25.35%)	12 (16.90%)	41 (57.75%)
Chi-square GOF	$\chi^2(2, n=71) = 19.80, p < .001, V = .37$		

Pairwise comparisons:

Comparison	Agree vs neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=30) = 1.2$, $p = .273$ (ns)	$\chi^2(1, n=59) = 9.00$, $p = .003, V = .39$	$\chi^2(1, n=53) = 15.87$, $p < .001, V = .55$

[1] Two participants asked in the comments to have their answers removed from the analysis of this question, hence the lower number of participants.

Small generative group:

Answer	Agree	Neither	Disagree
Frequency	15 (26.79%)	6 (10.71%)	35 (62.50%)
Chi-square GOF	$\chi^2(2, n=56) = 23.61, p < .001, V = .46$		

Pairwise comparisons:

Comparison	Agree vs neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=21) = 3.86$, $p = .050$ (ns)	$\chi^2(1, n=50) = 8$, $p = .005, V = .4$	$\chi^2(1, n=41) = 20.51, p < .001, V = .71$

The subject matter of the study of grammar:

All generative participants:

Answer	Patterns	Capacity	Other
Frequency	26 (35.62%)	33 (45.21%)	14 (19.18%)
Chi-square GOF	$\chi^2(2, n=73) = 7.50, p = .022, V = .23$		

Pairwise comparisons:

Comparison	Patterns vs Capacity	Patterns vs Other	Capacity vs Other
Chi-square GOF	$\chi^2(1, n=59) = .83$, $p = .362$ (ns)	$\chi^2(1, n=40) = 3.6$, $p= .058$ (ns)	$\chi^2(1, n=47) = 7.68$, $p = .006, V= .40$

Small generative group:

Answer	Patterns	Capacity	Other
Frequency	20 (35.09%)	28 (49.12%)	9 (15.79%)
Chi-square GOF	$\chi^2(2, n=57) = 9.58, p = .008, V = .29$		

Pairwise comparisons:

Comparison	Patterns vs Capacity	Patterns vs Other	Capacity vs Other
Chi-square GOF	χ^2(1, n=48) = 1.33, p = .248 (ns)	χ^2(1, n=29) = 4.17, p = .041 (ns)	χ^2(1, n=37) = 9.76, p = .002, V = .51

Structure rules and mental rules:

<u>All generative participants:</u>

Answer	Good hypothesis	As if	Other	Nothing
Frequency	21 (28.77%)	44 (60.27%)	3 (4.11%)	5 (6.85%)
Chi-square GOF	χ^2(3, n=73) = 59.11, p < .001, V = .52			

Pairwise comparisons:

Comparison	Good hypothesis vs As if	Good hypothesis vs Other	Good hypothesis vs Nothing
Chi-square GOF	χ^2(1, n=65) = 8.14, p = .004, V = .35	χ^2(1, n=24) = 13.5, p < .001, V = 0.75	χ^2(1, n=26) = 9.85, p = .002, V = .62

Comparison	As if vs Other	As if vs Nothing	Other vs Nothing
Chi-square GOF	χ^2(1, n=47) = 35.77, p < .001, V = .87	χ^2(1, n=49) = 31.04, p < .001, V = .80	χ^2(1, n=8) = 0.5, p = .480 (ns)

<u>Small generative group:</u>

Answer	Good hypothesis	As if	Other	Nothing
Frequency	17 (29.82%)	35 (61.40%)	3 (5.26%)	2 (3.51%)
Chi-square GOF	χ^2(3, n=57) = 50.16, p < .001, V = .54			

Pairwise comparisons:

Comparison	Good hypothesis vs As if	Good hypothesis vs Other	Good hypothesis vs Nothing
Chi-square GOF	χ^2(1, n=52) = 6.23, p = .013 (ns)	χ^2(1, n=20) = 9.8, p = .002, V = 0.7	χ^2(1, n=19) = 11.84, p < .001, V = .79

Comparison	As if vs Other	As if vs Nothing	Other vs Nothing
Chi-square GOF	$\chi^2(1, n=38) = 26.95$, $p < .001$, $V = .84$	$\chi^2(1, n=37) = 29.43$, $p < .001$, $V = .89$	$\chi^2(1, n=5) = 0.2$, $p = .655$ (ns)

The form of the implementation of mental rules:

<u>All generative participants:</u>

Answer	Represented	Not represented	Other
Frequency	12 (50.00%)	10 (41.67%)	2 (8.33%)
Chi-square GOF	$\chi^2(2, n=24) = 7$, $p = .030$, $V = .38$		

Pairwise comparisons:

Comparison	Represented vs Not represented	Represented vs Other	Not represented vs Other
Chi-square GOF	$\chi^2(1, n=22) = .18$, $p = .670$ (ns)	$\chi^2(1, n=14) = 7.14$, $p = .008$, $V = .71$	$\chi^2(1, n=12) = 5.33$, $p = .021$ (ns)

<u>Small generative group:</u>

Answer	Represented	Not represented	Other
Frequency	10 (50.00%)	9 (45.00%)	1 (5.00%)
Chi-square GOF	$\chi^2(2, n=20) = 7.3$, $p = .026$, $V = .43$		

Pairwise comparisons:

Comparison	Represented vs Not represented	Represented vs Other	Not represented vs Other
Chi-square GOF	$\chi^2(1, n=19) = .05$, $p = .819$ (ns)	$\chi^2(1, n=11) = 7.36$, $p = .007$, $V = .82$	$\chi^2(1, n=10) = 6.4$, $p = .011$, $V = .8$

B.2 Chi-square tests: generative participants, Experimental Syntax

Experts and ordinary speakers:

All generative participants:

Answer	Better	Equally good	Worse	Other
Frequency	17 (23.29%)	28 (38.36%)	20 (27.40%)	8 (10.96%)
Chi-square GOF	χ^2(3, n=73) = 11.22, p = .011, V = .23			

Pairwise comparisons:

Comparison	Better vs Equally good	Better vs Worse	Better vs Other
Chi-square GOF	χ^2(1, n=45) = 2.69, p = .101 (ns)	χ^2(1, n=37) = 0.24, p = .622 (ns)	χ^2(1, n=25) = 3.24, p = .072 (ns)

Comparison	Equally good vs Worse	Equally good vs Other	Worse vs Other
Chi-square GOF	χ^2(1, n=48) = 1.33, p = 0.248 (ns)	χ^2(1, n=36) = 11.11, p < .001, V = .56	χ^2(1, n=28) = 5.14, p = .023 (ns)

Small generative group:

Answer	Better	Equally good	Worse	Other
Frequency	13 (22.81%)	22 (38.60%)	14 (24.56%)	8 (14.04%)
Chi-square GOF	χ^2(3, n=57) = 7.07, p = .070 (ns)			

Data and theoretical virtues:

All generative participants:

Answer	Agre	Neither	Disagree
Frequency	54 (73.97%)	14 (19.18%)	5 (6.85%)
Chi-square GOF	χ^2(2, n=73) = 55.92, p < .001, V = .62		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	χ^2(1, n=68) = 23.53, p < .001, V = .59	χ^2(1, n=59) = 40.70, p < .001, V = .83	χ^2(1, n=19) = 4.26, p = .039 (ns)

Small generative group:

Answer	Agree	Neither	Disagree
Frequency	42 (73.68%)	11 (19.30%)	4 (7.02%)
Chi-square GOF	χ^2(2, n=57) = 43.05, p < .001, V = .61		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	χ^2(1, n =53) = 18.13, p < .001, V = .58	χ^2(1, n=46) = 31.39, p < .001, V = .83	χ^2(1, n=15) = 3.27, p = .071 (ns)

Gradience in grammar:

All generative participants:

Answers	Real degrees	Extra-grammatical	Other
Frequency	36 (49.32%)	22 (30.14%)	15 (20.55%)
Chi-square GOF	χ^2(2, n=73) = 9.40, p = .009, V = .25		

Pairwise comparisons:

Comparison	Other vs Extra-gramatical	Other vs Real	Extra-grammatical vs Real
Chi-square GOF	$\chi^2(1, n=24) = 1.32$, $p = 0.250$ (ns)	$\chi^2(1, n=38) = 8.65$, $p = .003$, $V = .41$	$\chi^2(1, n=58) = 3.38$, $p = .066$ (ns)

Small generative group:

Answer	Real degrees	Extra-grammatical	Other
Frequency	29 (50.88%)	16 (28.07%)	12 (21.05%)
Chi-square GOF	$\chi^2(2, n=57) = 8.32$, $p = .016$, $V = .27$		

Pairwise comparisons:

Comparison	Other vs Extra-grammatical	Other vs Real	Extra-grammatical vs Real
Chi-square GOF	$\chi^2(1, n=28) = 0.57$, $p = .450$ (ns)	$\chi^2(1, n=41) = 7.05$, $p = .008$, $V = .41$	$\chi^2(1, n=45) = 3.76$, $p = .053$ (ns)

Traditional methods:

All generative participants:

Answer	Agree	Neither	Disagree
Frequency	48 (65.75%)	7 (9.59%)	18 (24.66%)
Chi-square GOF	$\chi^2(2, n=73) = 37.01$, $p < .001$, $V = .50$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=55) = 30.56$, $p < .001$, $V = .75$	$\chi^2(1, n=66) = 13.64$, $p < .001$, $V = .45$	$\chi^2(1, n=25) = 4.84$, $p = .028$ (ns)

Small generative group:

Comparison	Agree	Neither	Disagree
Frequency	38 (66.67%)	7 (12.28%)	12 (21.05%)
Chi-square GOF	$\chi^2(2, n=57) = 29.16, p < .001, V = .51$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=45) = 21.36$, $p < .001, V = .69$	$\chi^2(1, n=50) = 13.52$, $p < .001, V = .52$	$\chi^2(1, n=19) = 1.32$, $p = .251$ (ns)

Experimental methods:

All generative participants:

Answer	Agree	Neither	Disagree
Frequency	45 (61.64%)	16 (21.92%)	12 (16.43%)
Chi-square GOF	$\chi^2(2, n=73) = 26.66, p < .001, V = .43$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=61) = 13.79$, $p < .001, V = .48$	$\chi^2(1, n=57) = 19.11$, $p < .001, V = .58$	$\chi^2(1, n=28) = .57$, $p = .450$ (ns)

Small generative group:

Answer	Agree	Neither	Disagree
Frequency	33 (57.89%)	14 (24.56%)	10 (17.54%)
Chi-square GOF	$\chi^2(2, n=57) = 15.90, p < .001, V = .37$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=47) = 7.68$, $p = .006$, $V = .40$	$\chi^2(1, n=43) = 12.30$, $p < .001$, $V = .53$	$\chi^2(1, n=24) = .67$, $p = .414$ (ns)

B.3 Chi-square tests: generative participants, other questions

Syntactic intuitions and other types of linguistic evidence:

All generative participants:

Answer	In some cases	Never stand alone	Other
Frequency	64 (87.67%)	8 (10.96%)	1 (1.37%)
Chi-square GOF	$\chi^2(2, n=73) = 98, p < .001, V = .82$		

Pairwise comparisons:

Comparison	In some cases vs Never stand alone	In some cases vs Other	Never stand alone vs Other
Chi-square GOF	$\chi^2(1, n=72) = 43.56$, $p < .001, V = .78$	$\chi^2(1, n=65) = 61.06$, $p < .001, V = .97$	$\chi^2(1, n=9) = 5.44$, $p = .020$ (ns)

Small generative group:

Answer	In some cases	Never stand alone	Other
Frequency	51 (89.47%)	5 (8.77%)	1 (1.75%)
Chi-square GOF	$\chi^2(2, n=57) = 81.26, p < .001, V = .84$		

Pairwise comparisons:

Comparison	In some cases vs Never stand alone	In some cases vs Other	Never stand alone vs Other
Chi-square GOF	$\chi^2(1, n=56) = 37.79$, $p < .001, V = .82$	$\chi^2(1, n = 52) = 48.08$, $p < .001, V = .96$	$\chi^2(1, n=6) = 2.67$, $p = .103$ (ns)

Significance (1):

All generative participants:

Answer	Yes	No
Frequency	70 (95.89%)	3 (4.11%)
Chi-square GOF	$\chi^2(1, n=73) = 61.49, p < .001, V = .92$	

Small generative group:

Answer	Yes	No
Frequency	54 (94.72%)	3 (5.26%)
Chi-square GOF	$\chi^2(1, n=57) = 45.63, p < .001, V = .89$	

Significance (2):

All generative participants:

Answer	Agree	Neither	Disagree
Frequency	65 (89.04%)	3 (4.11%)	5 (6.85%)
Chi-square GOF	$\chi^2(2, n=73) = 102, p < .001, V = .84$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=68) = 56.53$, $p < .001, V = .91$	$\chi^2(1, n=70) = 51.43$, $p < .001, V = .86$	$\chi^2(1, n=8) = 0.5$, $p = .480$ (ns)

Small generative group:

Answer	Agree	Neither	Disagree
Frequency	51 (89.47%)	2 (3.51%)	4 (7.02%)
Chi-square GOF	$\chi^2(2, n=57) = 80.95, p < .001, V = .84$		

Pairwise comparisons:

Comparison	Agree vs neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n = 53) = 45.30$, $p < .001$, $V = .92$	$\chi^2(1, n = 55) = 40.16$, $p < .001$, $V = .85$	$\chi^2(1, n = 6) = 0.67$, $p = .414$ (ns)

B.4 Chi-square tests: non-generative participants, the etiology debate

The origin of syntactic intuitions:

Answer	Reflections	Competence	Other
Frequency	40 (65.57%)	14 (22.95%)	7 (11.48%)
Chi-square GOF	$\chi^2(2, n=61) = 29.74, p < .001, V = .49$		

Pairwise comparisons:

Comparison	Reflections vs Competence	Reflections vs Other	Competence vs Other
Chi-square GOF	$\chi^2(1, n=54) = 12.52$, $p < .001, V = .48$	$\chi^2(1, n=47) = 23.17$, $p < .001, V = .70$	$\chi^2(1, n=21) = 2.33$, $p = .127$ (ns)

Acceptability and grammaticality:

Answer	Acceptability	Grammaticality	Other
Frequency	36 (59.02%)	9 (14.75%)	16 (26.23%)
Chi-square GOF	$\chi^2(2, n=61) = 19.31, p < .001, V = .40$		

Pairwise comparisons:

Comparison	Acceptability vs Grammaticality	Acceptability vs Other	Grammaticality vs Other
Chi-square GOF	$\chi^2(1, n=45) = 16.2$, $p < .001, V = .6$	$\chi^2(1, n=52) = 7.69$, $p = .006, V = .38$	$\chi^2(1, n=25) = 1.96$, $p = .162$ (ns)

The role of mental grammar:

Answer	Poor description	Not literal	Other	Actual process
Frequency	28 (45.90%)	29 (47.54%)	3 (4.92%)	1 (1.64%)
Chi-square GOF	$\chi^2(3, n=61) = 46.21, p < .001, V = .50$			

Pairwise comparisons:

Comparison	Poor description vs Not literal	Poor description vs Other	Poor description vs Actual process
Chi-square GOF	$\chi^2(1, n=57) = 0.02$, $p = .895$ (ns)	$\chi^2(1, n=31) = 20.16$, $p < .001$, $V = .81$	$\chi^2(1, n=29) = 25.14$, $p < .001$, $V = .93$

Comparison	Not literal vs Other	Not literal vs Actual process	Other vs Actual process
Chi-square GOF	$\chi^2(1, n=32) = 21.13$, $p < .001$, $V = .81$	$\chi^2(1, n=30) = 26.13$, $p < .001$, $V = .93$	$\chi^2(1, n=4) = 1$, $p = .317$ (ns)

Fallibility of intuitions:

Answer	Agree	Neither	Disagree
Frequency	29 (47.54%)	22 (36.07%)	10 (16.39%)
Chi-square GOF	$\chi^2(2, n=61) = 9.08$, $p = .011$, $V = .27$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=51) = 0.96$, $p = .327$ (ns)	$\chi^2(1, n=39) = 9.26$, $p = .002$, $V = .49$	$\chi^2(1, n=32) = 4.5$, $p = .034$ (ns)

The subject matter of the study of grammar:

Answer	Patterns	Capacity	Other
Frequency	42 (68.85%)	10 (16.39%)	9 (14.75%)
Chi-square GOF	$\chi^2(2, n=61) = 34.66$, $p < .001$, $V = .53$		

Pairwise comparisons:

Comparison	Patterns vs Capacity	Patterns vs Other	Capacity vs Other
Chi-square GOF	$\chi^2(1, n=52) = 19.69$, $p < .001$, $V = .62$	$\chi^2(1, n=51) = 21.35$, $p < .001$, $V = .65$	$\chi^2(1, n=19) = 0.05$, $p = .819$ (ns)

Structure rules and mental rules:

Answer	Good hypothesis	As if	Other	Nothing
Frequency	5 (8.20%)	34 (55.74%)	5 (8.20%)	17 (27.87%)
Chi-square GOF	$\chi^2(3, n=61) = 37.03, p < .001, V = .45$			

Pairwise comparisons:

Comparison	Good hypothesis vs As if	Good hypothesis vs Other	Good hypothesis vs Nothing
Chi-square GOF	$\chi^2(1, n=39) = 21.56$, $p < .001$, $V = .74$	Not compared	$\chi^2(1, n=22) = 6.55$, $p = .011$ (ns)

Comparison	As if vs Other	As if vs Nothing	Other vs Nothing
Chi-square GOF	$\chi^2(1, n=39) = 21.56$, $p < .001$, $V = .74$	$\chi^2(1, n=51) = 5.67$, $p = .017$ (ns)	$\chi^2(1, n=22) = 6.55$, $p = .011$ (ns)

The form of the implementation of mental rules:

Answer	Represented	Not represented	Other
Frequency	2 (20%)	4 (40%)	4 (40%)
Chi-square GOF	$\chi^2(2, n=10) = 0.8$, $p = .670$ (ns)		

B.5 Chi-square tests: non-generative participants, Experimental Syntax

Experts and ordinary speakers:

Answer	Better	Equally good	Worse	Other
Frequency	10 (16.39%)	21 (34.43%)	25 (40.98%)	5 (8.20%)
Chi-square GOF	χ^2(3, n=61) = 17.10, p < .001, V = .31			

Pairwise comparisons:

Comparison	Better vs Equally good	Better vs Worse	Better vs Other
Chi-square GOF	χ^2(1, n=31) = 3.90, p = .048 (ns)	χ^2(1, n=35) = 6.43, p = .011 (ns)	χ^2(1, n=15) = 1.67, p = .197 (ns)

Comparison	Equally good vs Worse	Equally good vs Other	Worse vs Other
Chi-square GOF	χ^2(1, n=46) = 0.35, p = .555 (ns)	χ^2(1, n=26) = 9.85, p = .002, V = .62	χ^2(1, n=30) = 13.33, p < .001, V = .67

Data and theoretical virtues:

Answer	Agree	Neither	Disagree
Frequency	55 (90.16%)	3 (4.92%)	3 (4.92%)
Chi-square GOF	χ^2(2, n=61) = 88.66, p < .001, V = .85		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	χ^2(1, n=58) = 46.62, P < .001, V = .90	χ^2(1, n=58) = 46.62, P < .001, V = .90	Not compared

Gradience in grammar:

Answer	Other	Extra-grammatical	Real degrees
Frequency	7 (11.48%)	13 (21.31%)	41 (67.21%)
Chi-square GOF	$\chi^2(2, n=61) = 32.39, p < .001, V = .52$		

Pairwise comparisons:

Comparison	Other vs Extra-grammatical	Other vs Real degrees	Extra-grammatical vs Real degrees
Chi-square GOF	$\chi^2(1, n=20) = 1.8$, $p = .180$ (ns)	$\chi^2(1, n=48) = 24.08$, $p < .001, V = .71$	$\chi^2(1, n=54)= 14.52$, $p < .001, V = .52$

Traditional methods:

Answer	Agree	Neither	Disagree
Frequency	17 (27.87%)	12 (19.67%)	32 (52.46%)
Chi-square GOF	$\chi^2(2, n=61) = 10.66, p = .005, V = .30$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=29) = 0.86$, $p = .353$ (ns)	$\chi^2(1, n=49) = 4.59$, $p = .032$ (ns)	$\chi^2(1, n=44) = 9.09$, $p = .003, V = .45$

Experimental methods:

Answer	Agree	Neither	Disagree
Frequency	46 (75.41%)	10 (16.39%)	5 (8.20%)
Chi-square GOF	$\chi^2(2, n=61) = 49.21, p < .001, V = .64$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=56) = 23.14$, $p < .001, V = .64$	$\chi^2(1, n=51) = 32.96$, $p < .001, V = .80$	$\chi^2(1, n=15) = 1.67$, $p = .197$ (ns)

B.6 Chi-square tests: non-generative participants, other questions

Syntactic intuitions and other types of linguistic evidence:

Answer	In some cases	Never stand alone	Other
Frequency	27 (44.26%)	33 (54.10%)	1 (1.64%)
Chi-square GOF	$\chi^2(2, n=61) = 28.46, p < .001, V = .48$		

Pairwise comparisons:

Comparison	In some cases vs Never stand alone	In some cases vs Other	Never stand alone vs Other
Chi-square GOF	$\chi^2(1, n=60) = 0.6$, $p = .439$ (ns)	$\chi^2(1, n=28) = 24.14$, $p < .001, V = .93$	$\chi^2(1, n=34) = 30.12$, $p < .001, V = .94$

Significance (1):

Answer	Yes	No
Frequency	55 (90.16%)	6 (9.84%)
Chi-square GOF	$\chi^2(1, n=61) = 39.36, p < .001, V = .80$	

Significance (2):

Answer	Agree	Neither	Disagree
Frequency	54 (88.52%)	4 (6.56%)	3 (4.92%)
Chi-square GOF	$\chi^2(2, n=61) = 83.64, p < .001, V = .83$		

Pairwise comparisons:

Comparison	Agree vs Neither	Agree vs Disagree	Neither vs Disagree
Chi-square GOF	$\chi^2(1, n=58) = 43.10$, $p < .001, V = .86$	$\chi^2(1, n=57) = 45.63$, $p < .001, V = .89$	$\chi^2(1, n=7) = 0.14$, $p = .706$ (ns)

Appendix C

Content analyses

Overview tables for all the content analyses performed for this dissertation are presented on the following pages. The content analysis method is described in section 6.1.1.

Analyses are presented for both the alternative answers participants provided ("Supply other answer") and for the optional comments. The questions are presented in the same order here as in section 6.2. First, the generative participants' comments on questions relating to the etiology debate are presented, then their comments on questions relating to the Experimental Syntax debate, and finally their comments on other questions. The non-generative participants' comments are then presented in the same order.

C.1 Content analyses: generative participants, the etiology debate

The origin of syntactic intuitions:

Other, n = 16:

Theme	*Example*	*No.*	*%*
Both	can be both	7	44%
Competence	they allow us to infer what the speaker's competence is	3	19%
Other	They reflect processing complexity, which could be labelled 'competence', but needn't be.	3	19%
Neither	Neither	1	6%
Comment on question	I do not understand the question	1	6%
Uncategorised	see below	1	6%
Total		16	100%

Optional, n = 10:

Theme	*Example*	*No.*	*%*
Competence	[...] for sure it reflects something about the speaker’s competence (defined in a very broad sense).	3	30%
Other	They are reliable because they are systematic. That is sufficient.	3	30%
Process	What we should take a given intuition to express depends at least in part on the task that was used to elicit the intuition.	2	20%
Not reflection	Option A is impossible: an intuition is never a reflection, it's just spontaneous and unreflected.	1	10%
Comment on question	[...] I find your simply binarity here misleading.	1	10%
Total		10	100%

Acceptability and grammaticality:

Other, n = 25:

Theme	*Example*	*No.*	*%*
Both	Both, distinctly	15	60%
Focus on process	robust across different lexicalizations and contexts.	3	12%
Acceptability	It is not possible to directly assess grammaticality intuitions; all judgements are acceptability judgements	2	8%
Other	felicity within a context	2	8%
Distinction	I do not share the distinction.	2	8%
Neither	I really see both as problematic.	1	4%
Total		25	100%

Optional, n = 23:

Theme	*Example*	*No.*	*%*
Not grammaticality	there is no such thing as a grammaticality intuition. grammaticality is a theoretical concept; acceptability is an empirical one.	7	30%
Distinction	It is not easy to draw a clear distinction between the two. [...]	4	17%
Focus on process	[...] This exx from English would need some context and appropriate prosody to accurately judged as grammatical.	3	13%
Other	The analyst can use all kinds of speaker judgments as evidence.	3	13%
Link	We can't access linguistic processes directly... in this regard, it's really difficult to tease apart 'acceptability' and 'grammaticality'	2	9%
Grammaticality, elaboration	But the theory will need to provide evidence-based arguments as to why such sentences are generally not accepted as well-formed by the speakers	2	9%
Comment on question	I don't think native speakers would find it 'ungrammatical' - I don't think they would use that term.	2	9%
Total		23	100%

The role of mental grammar:

Other, n = 8:

Theme	*Example*	*No.*	*%*
And other factors	this is the ideal situation, but many elements can influence the judgments (attitudes, prescriptivism, ...)	4	50%
Comment on question	It is unclear what you mean by the statement above.	2	25%
Depends	I think it will vary depending on the construction type under investigation.	1	12.5%
Other	We deduce the mental grammar from the intuitions, I'd say.	1	12.5%
Total		8	100%

Optional, n = 8:

Theme	*Example*	*No.*	*%*
And other factors	there are other factors that affect intuitions other than the mental grammar, so the intuitions are not strictly 'deduced'	3	37.5%
Poor description, elaborate	Doesn't seem like a useful description	3	37.5%
Actual process, elaborate	With the caveat that this is not a conscious process in most cases.	1	12.5%
Good way to talk, elaborate	This way of talking gets the conversation going.	1	12.5%
Total		8	100%

Fallibility of intuitions:

Optional, n = 20:

Theme	*Example*	*No.*	*%*
“Performance factors”	Many non-performance related issue can confound grammaticality judgements (semantics, prosody, dialectal differences etc.).	7	35%
Question unclear	This question is unclear to me, you should disregard my answer.	4	20%
Agree, elaborate	Linguistic illusions (work by Colin Phillips and many of his students) are a very good example of cases where speakers are 'mistaken' about the grammatical properties of sentences.	4	20%
“Mistaken”	Acceptability intuitions are a scalar thing. To be 'mistaken' is not a concept that I can meaningfully apply to intuitions, even if I imagine that performance is stripped off.	3	15%
Comment on question	This is a really interesting question!	2	10%
Total		20	100%

The subject matter of the study of grammar:

Other, n = 14:

Theme	*Example*	*No.*	*%*
Both	Both of the above	9	64%
Both and other	both, plus how these phenomena are acquired	2	14%
One by means of the other	The linguistic capacity of the mind by means of behavioural observations; the latter comes from the minds of individuals.	2	14%
Other	the processing mechanism(s) of language as they unfold in real time	1	7%
Total		14	100%

Optional, n = 15:

Theme	*Example*	*No.*	*%*
Both	I would have liked to answer 'both'	7	47%
Patterns and other	I want to find patterns and also learn something about the choices speakers make.	2	13%
Other, elaboration	Behaviour is not the end of the story, but we gather psycho-social info (unless doing implicit or neuro measures of language processing)	2	13%
Comment on question	Neither are the two given alternatives exhaustive, nor are they (mutually) exclusive.	2	13%
Capacity, elaboration	[...] I am working on a particular causal power which seems particularly relevant to language, which you have called 'the linguistic capacity of the mind'.	1	7%
Uncategorised	'linguistic capacity of the mind' is a bit vague for me.	1	7%
Total		15	100%

Structure rules and mental rules:

Other, n = 3:

Theme	*Example*	*No.*	*%*
Not processing	From the structure rules we observe, we can infer that the mind has a computational system that implements these rules (e.g. a competence), but we cannot infer anything about how these rules are processed online by speakers.	2	75%
Comment on question	These all say the same thing	1	25%
Total		3	100%

Optional, n = 13:

Theme	*Example*	*No.*	*%*
Good hypothe-sis, elaborate	Although option B makes a lot of sense too. Option A indeed seems to offer a better (more direct) hypothesis.	3	23%
As if, elaborate	It is just like any other science.	3	23%
Levels	[...] The 'levels hypothesis' (Marr, Pylyshyn) is the best way we have of thinking of this.	2	15%
Comment on question	It's not clear what 'implemented in the minds of speakers' means here.	2	15%
Nothing, elaborate	Neural matter might have another structure than its products.	1	8%
Not processing	NB: competence rules aren't about *processing* language, are they...?	1	8%
Uncategorised	You should separate mind/brain in these descriptions. Regardless, because the Generative perspective is immune to criticism from 'just performance data' it is also unable to inform us about processing in any way. [...]	1	8%
Total		13	100%

The form of the implementation of mental rules:

Other, n = 2:

Theme	*Comment*	*No.*	*%*
Comment on question	Surely the workings of language take place in the mind. However, if we talk about 'implementation' we should talk about the brain.	1	50%
Levels	The rules of grammar are a high level description of certain regularities in computational processes	1	50%
Total		2	100%

Optional, n = 4:

Theme	*Example*	*No.*	*%*
Not represented, elaborate	It would be very naive to assume that the rules of grammar are explicitly represented in the mind, whatever that actually means.	2	50%
Represented, elaborate	But knowing what to look for would be very hard	1	25%
Levels	Grammatical rules are implemented in the mind the way the text of a high-level program is related to its execution after being compiled into assembly language.	1	25%
Total		4	100%

C.2 Content analyses: generative participants, Experimental Syntax

Experts and ordinary speakers:

Other, n = 8:

Theme	*Example*	*No.*	*%*
Depends	i think it depends on the speaker, the construction type, and how the judgments are elicited.	7	87.5%
Don't know	I don't know	1	12.5%
Total		8	100%

Optional, n = 19:

Theme	*Example*	*No.*	*%*
Depends	Linguists know to abstract away from things we think we are not interested in (like style, or frequency), but they may have theory-clouded judgments.	6	32%
Experts worse	It is known that repeated exposure to a marginally ill-formed structure weakens judgments.	5	26%
Experts better	Experts are generally good at recognizing confounds. That's what they're trained to do.	5	26%
Focus on process	The naive native speakers have to be given the right instructions, but they are probably easier to train in the correct methods. [...]	2	11%
Comment on question	The above response assumes, again, we're dealing with acceptability and not grammaticality judgments	1	5%
Total		19	100%

Data and theoretical virtues:

Optional, n = 17:

Theme	*Example*	*No.*	*%*
Agree, elaborate	It doesn't matter how elegant a theory is if it accounts for the wrong data.	7	41%
Both	One always needs to balance compelling facts and compelling generalizations that don't capture all known facts. Welcome to science!	7	41%
Theoretical virtues important	I suppose that theories can be useful for different reasons.	3	18%
Total		17	100%

Gradience in grammar:

Other, n = 15:

Theme	*Example*	*No.*	*%*
No gradience	There is no such thing as a 'degree of grammaticality'	3	20%
Both	A bit of both	3	20%
Uncategorised	They can certainly help.	3	20%
Don't know	Unsure	2	13%
Comment on question	difficult question :)	2	13%
Open question	I don't think we know yet whether gradience in judgments reflects real gradience or other effects. [...]	1	7%
Assumption	The assumption that everything sits on a single dimension of 'grammaticality' is likely a fiction.	1	7%
Total		15	100%

Optional, n = 8:

Theme	*Example*	*No.*	*%*
Open question	still an open question; it is still not at all clear whether grammaticality is gradient or binary	3	37.5%
No gradience in grammar	Grammaticality IS binary. [...]	2	25%
Focus on process	They must be very 'well-designed'.	2	25%
Comment on question	It's very difficult to say what 'grammaticality' means in abstraction from a particular theory of grammar.	1	12.5%
Total		8	100%

Traditional methods:

Optional, n = 20:

Theme	*Example*	*No.*	*%*
Starting point	Of course, this should only be the starting point.	3	15%
If uncontroversial	It usually produces robust and repeatable data, but there are often very subtle judgments that are not stable. [...]	3	15%
Focus on process	Assuming best practices are followed [...]	3	15%
Disagree, elaborate	In the long run, intuitions of individuals are just not enough. [...]	3	15%
Agree, elaborate	The proof is in decades of reliable findings. [...]	2	10%
Group reinforcement	I see this leading to positive reinforcement of marginal analyses sometimes. Crony-ism	2	10%
Idiolects	Consulting one person is always good evidence of their own grammar [...]	1	5%
Many sources	It is good to have data points from as many sources as possible for control purposes.	1	5%
Grain of salt	but 'cum grano salis'	1	5%
Comment on question	Could I just comment that I think the phrasing 'evidence for theories grammar' is a bit off. [...]	1	5%
Total		20	100%

Experimental methods:

Optional, n = 26:

Theme	*Example*	*No.*	*%*
Both	I tend to work this way, but it is not the only way and the different methods are mutually informative. [...]	9	35%
If controversial	If the data are contentious, then it may be necessary.	7	27%
Idiolects	Grammars vary and a small group of speakers could be enough.	3	11%
Uncategorised	There is a literature about this.	3	11%
If feasible	if possible, yes, but in the case of endangered languages we work with what we have.	2	8%
Disagree, elaborate	I use those methods all the time. They rarely yield more than we get from simply asking a couple of friends.	1	4%
Agree, elaborate	[...] For two reasons: 1) this supplements the data, and more data is always better, and 2) it shows that the conclusions are representative. [...]	1	4%
Total		26	100%

C.3 Content analyses: generative participants, other questions

Syntactic intuitions and other types of linguistic evidence:

Other, n = 1:

Theme	*Example*	*No.*	*%*
Can stand alone	Can always stand alone as evidence	1	100%
Total		1	100%

Optional, n = 20:

Theme	*Example*	*No.*	*%*
Can stand alone, elaborate	If they're no evidence, then we have virtually null information from all the grammar books based on intuitions. A highly unlikely idea.	7	35%
Focus on process	Context must be given with the test sentence	4	20%
Depends	It depends what kind of evidence you want - individual grammar or community info. [...]	3	15%
Not alone	... but should not be used alone.	3	15%
Other factors	Certain empirical phenomena is intrinsically connected to speakers judgments, e.g., ambiguity.	2	10%
Comment on question	It is not specified here _for what_ syntactic intuitions might constitute evidence. [...]	1	5%
Total		20	100%

Significance (1):

Other, n = 1: recoded (see section 6.1.4.1)

Optional, n = 11:

Theme	*Example*	*No.*	*%*
When	in my own experiments on the topic :)	7	64%
Why	I have become worried that speakers can use extra grammatical reasoning, for instance mathematical reasoning, on things like quantifier scope.	3	27%
Significance	If anyone claims to be a linguist and answers 'no' to the above question then one must question whether they are, in fact, a linguist.	1	9%
Total		11	100%

Significance (2):

Optional, n = 12:

Theme	*Example*	*No.*	*%*
Agree, elaborate	In the sense that all scientists should always take care and be reflective about methodology.	9	75%
Depends	It totally depends on the phenomenon one wants to analyze.	2	17%
Disagree, elaborate	Intuitions can tell us something about grammars, that's true. In my opinion, the question is what, not why. [...]	1	8%
Total		12	100%

Any other comments:

Optional, n = 17

Theme	*Example*	*No.*	*%*
Comment on questionnaire	I find that some of the given options suggest a binarity that just doesn't exist. [...]	9	53%
Central point	If native speakers' intuitive judgements cannot be used as evidence in linguistics, what at all can be?	6	35%
Significance	I think that linguists do not typically think about these issues	2	12%
Total		17	100%

C.4 Content analyses: non-generative participants, the etiology debate

The origin of syntactic intuitions:

Other, n = 7:

Theme	*Example*	*No.*	*%*
Both	Both can be true. It depends on the examined phenomenon, the speaker's sociolinguistic state or his/her apability to distinguish form and meaning etc.	2	29%
Other	They are merely indications of perceptions and/or preferences.	2	29%
Assumption of question	They are not reliable	1	14%
Comment on question	Again, both options are not mutually exclusive	1	14%
Uncategorised	speakers are to some degree reliable concerning judgements of acceptability.	1	14%
Total		7	100%

Optional, n = 9:

Theme	*Example*	*No.*	*%*
“Reflection”	while intuitions (first impressions) are not really reflections (thinking about it)	2	22%
Assumption of question	They are not reliable as evidence, though intuitions can show patterns like all other human behaviour.	2	22%
And other	… including competence beyond primary (form-to-meaning) semiotics (e.g., pragmatic competence, 'literary competence')	1	11%
Depends	A lot depends on who the 'speakers' are that we're talking about.	1	11%
Elaboration	I’ll grudgingly accept the first option. Of course speakers are to some degree good judges of acceptable usage	1	11%
Grain of salt	but speakers' reflections on language should always be taken with a pinch of salt: look at how people think about how the outside world works (many people believe the Earth is flat, even in countries with a good education system)	1	11%
Comment on question	Beginning to think this survey just doesn't work for linguists who see language competence and performance as two sides of the same coin (i.e., usage-based)	1	11%
Total		9	100%

Acceptability and grammaticality:

Other, n = 16:

Theme	*Example*	*No.*	*%*
Both	both	8	50%
Distinction	I do not accept the grammatical/acceptable distinction	2	13%
Acceptability	Speakers can say whether the sentence is OK or not, but if it is not, it is the linguists task to say wether it is ungrammatical or it is not OK because of other reasons.	1	6%
Assumption of question	intuitions aren't 'evidence' at all	1	6%
Focus on process	. . . concretely specified with regard to the procedure for eleciting the judgements.	1	6%
Neither	neither	1	6%
Other	replicable	1	6%
Comment on question	This statement is unclear...	1	6%
Total		16	100%

Optional, n = 13:

Theme	*Example*	*No.*	*%*
Comment on question	the question is not clear	3	23%
Other	... and others: e.g. familiarity intuitions can be very informative too	3	23%
Focus on process	[...] The RECORD of an explicit intuitive judgement must not be treated as a transparent reflection of a real intuition that exists independently of the elicitation procedure. [...]	2	15%
Distinction	[...] In fact, grammaticality should be defined purely by acceptable usage	2	15%
Acceptability	[...] acceptability is a better concept given the gradiency of language	2	15%
Both	Some people are able to give insights about both of these criteria.	1	8%
Total		13	100%

The role of mental grammar:

Other, n = 3:

Theme	*Example*	*No.*	*%*
Poor description	It's a poor description of how acceptability judgements are formed, but a good description of how intuitions about such things as what a sentence means are formed.	1	33%
Other	for me the constructionist approach seems more likely - when I reflect on my usage of L2 and L3, it looks like there really exist some chunks	1	33%
Uncategorised	This is a confused way of putting it	1	33%
Total		3	100%

Optional, n = 5:

Theme	*Example*	*No.*	*%*
Not deduced	We might be able to deduce mental grammar from intuitions, but we cannot deduce intuitions from mental grammar, as we have access only to intuitions.	2	40%
Other	I would say that intuitions are due to the interiorization of a social rule	1	20%
Uncategorised	This should be true of authors of grammar books	1	20%
Comment on question	Don't think this question (or the options) are very clear. [...]	1	20%
Total		5	100%

Fallibility of intuitions:

Optional, n = 18:

Theme	*Example*	*No.*	*%*
Performance central	[...] Performance factors are part of the grammar.	4	22%
Not possible	I don't think it is ever possible to strip away performance factors	3	17%
Comment on question	Strangely formulated statement	3	17%
Assumption of question	This statement assumes that 'the' grammatical properties of the sentence exist completely separate from (and invariant among different) speakers	3	17%
Distinction	The distinction between 'competence' and 'performance' is a false distinction.	2	11%
Depends	If we are dealing with clear cases, I would say, 'I disagree'. If we are dealing with less-than-clear cases (by far the majority), I would say that I agree.	1	6%
Agree, elaborate	[...] Since patterns are found across sentences, not within one sentence, a speaker have lots of wrong intuitions about an isolated sentence.	1	6%
Uncategorised	Some people are not competent users of any language.	1	6%
Total		18	100%

The subject matter of the study of grammar:

Other, n = 9:

Theme	*Example*	*No.*	*%*
Both	both of the above	6	67%
Other	the cognitive processes operating	2	22%
Comment on question	The above are not mutually exclusive.	1	11%
Total		9	100%

Optional, n = 5:

Theme	*Example*	*No.*	*%*
Not mental capacities	The mind cannot be observed directly; tree diagrammes do not represent the mind	2	40%
And other	...and how these influence interpretation in interaction	1	20%
Both	[...] What I'm interested in is a bit of both: understanding how 'big ideas' like 'linguistic capacity of the mind' can be made concrete and evaluated in systematic ways.	1	20%
One in order to study the other	Language is a neuro-cognitive capacity which has as overt manifestation a system of linguistics units: To understand Language we need to understand its manifestation(s).	1	20%
Total		5	100%

Structure rules and mental rules:

Other, n = 5:

Theme	*Example*	*No.*	*%*
Just hypothesis	[...] In principle, this hypothesis is acceptable as such, but to test it, psycholinguistic and brain research would be needed.	2	40%
Both	Both A and B. We can only infer that the mind works 'as if' it was following those rules --- but it is nonetheless a good working hypothesis that the rules are in fact implemented in the minds of speakers.	1	20%
Not rules	Rules may not be the best conceptualisation of the process	1	20%
Other	I'd say structure rules result from observing language; they can help monitoring language production when the speaker needs help	1	20%
Total		5	100%

Optional, n = 11:

Theme	*Example*	*No.*	*%*
Elaboration, nothing	[...] Correlating rules that linguists arrive at with cognitive reality is haphazardous.	4	36%
Comment on question	Poorly worded question. [...]	2	18%
Uncategorised	If forced to choose between the three options, that is the one I would choose.	2	18%
Elaboration, good hypothesis	Of course, a good hypothesis can still be falsified	1	9%
Just hypothesis	From the rules alone we can only infer the 'as if'. Based on that, we hypothesize an implementation. [...]	1	9%
Depends	The answer also depends on the definition of 'how the mind processes language'. [...]	1	9%
Total		11	100%

The form of the implementation of mental rules:

Other, n = 4:

Theme	*Example*	*No.*	*%*
Not rules	Rules may not be the best conceptualisation of this process.	1	25%
Not transparent	There must be an interface between whatever is in the mind andobservable patterns of grammar. But that doesn't mean the mind operates on what we observe. [...]	1	25%
Other	The rules of grammar may be implemented in the mind.	1	25%
Comment on question	Is the mind the same as the brain?	1	25%
Total		4	100%

Optional, n = 3:

Theme	*Example*	*No.*	*%*
Elaborate, not represented	I am doubtful that much of anything is 'explicitly represented in the mind', beyond perhaps direct sensory data. [...]	2	67%
Elaborate, represented	[...] Firstly it, unlike the two alternatives, instantly explains how come there seem to be rules. Secondly it explains why language is so simple and regular ('below the surface').	1	33%
Total		3	100%

C.5 Content analyses: non-generative participants, Experimental Syntax

Experts and ordinary speakers:

Other, n = 5:

Theme	*Example*	*No.*	*%*
Different	equally flawed, but differently.	3	60%
Depends	depends on how familiar the linguist is with the languages he/she is working with	1	20%
Uncategorised	neither worse nor better	1	20%
Total		5	100%

Optional, n = 20:

Theme	*Example*	*No.*	*%*
More confounds using experts	The linguist will be biased by his formal knowledge or beliefs regarding grammar...	10	50%
Depends	assuming their level of education is higher than that of an average person.	4	20%
Elaborate, experts better	[...] linguists are more accurate about languages than native speakers.	2	10%
Uncategorised	I am unsure about the theretical (and practical) relevance of the concept of native speaker.	2	10%
Should not be used alone	...but reliance on introspection alone is not good scientific practice	1	5%
Different	Not worse, not better, not equal, just different, reflecting different experience and exposure.	1	5%
Total		20	100%

Data and theoretical virtues:

Optional, n = 14:

Theme	*Example*	*No.*	*%*
Both	Hopefully, nothing prevents a theory from being both	4	29%
Depends	[...] I think this is to some extent a matter of taste, and to some extent a matter of what one intends to do with the theory, and there is no one right answer.	3	21%
Fruitfulness	Fruitful is fruitful.	3	21%
Elaborate, agree	As a functionalist, my primary aim is the description of language and its functions, and the building of theories is secondary	2	14%
Uncategorised	cf. Sextus Empiricus 'Adversos Grammaticos'	2	14%
Total		14	100%

Gradience in grammar:

Other, n = 7:

Theme	*Example*	*No.*	*%*
Both	The first answer is occasionally applicable. The second is far more usually applicable.	2	29%
Uncategorised	They are often the only way	2	29%
Depends	It can depend on the case and on the definition of grammaticality.	1	14%
Distinction	I do not accept the term 'grammaticality' as defined.	1	14%
Gradient	both are useful: binary and gradient scales yield gradient results if analysed properly; binary forced choices seem to be most sensitive	1	14%
Total		7	100%

Optional, n = 8:

Theme	*Example*	*No.*	*%*
Elaborate, gradience	Although it's true that they may not be perfect, they are closer to getting us the right answers we are looking for	3	37.5%
Not just grammaticality	There is no way to test 'just grammaticality' with intuitions. No scale will help. [...]	2	25%
Problems of gradient scales	Informants tend to have problems with a five-point scale. Continuous scales are haphazardous. But forcing a good/bad judgment isn't desirable either. [...]	2	25%
Comment on question	Although gradient does not mean 'degrees' (something that has degrees can still be categorical, and not gradient).	1	12.5%
Total		8	100%

Traditional methods:

Optional, n = 15:

Theme	*Example*	*No.*	*%*
Starting point	these intuitions can be a good starting point to do some serious corpus-derived and experimental work	7	47%
Depends	again, it is ok if we are dealing with clear cases. Otherwise, I would look for empirical data	4	27%
Elaborate, disagree	I do not accept intuitions as of much validity.	3	20%
Uncategorised	cf. my comment to the last question	1	7%
Total		15	100%

Experimental methods:

Optional, n = 18:

Theme	*Example*	*No.*	*%*
Elaborate, agree	Difficult and time consuming, but worth the while	5	28%
If feasible	If possible, this would be an ideal. It isn't always possible, though.	4	22%
Elaborate, neither	I don't think statistical tests need to be used, and I don't think 'large numbers of speakers' is needed, although there should be more than one or two.	3	17%
Depends	The 'should' is far too strong here. If we're talking about the order of determiner-noun across languages, I strongly disagree. If we're talking about long-distance wh-questions, then yes	3	17%
Focus on process	and using proper statistical procedure to ensure data is not biased	2	11%
Elaborate, disagree	I read the question as meaning 'should ONLY be'. The method described is of no value or interest to me, but that doesn't mean it is of no interest or value to scholarship more generally.	1	6%
Total		18	100%

C.6 Content analyses: non-generative participants, other questions

Syntactic intuitions and other types of linguistic evidence:

Other, n = 1:

Theme	*Example*	*No.*	*%*
Starting point	can offer hypotheses to be tested against evidence.	1	100%
Total		1	100%

Optional, n= 19:

Theme	*Example*	*No.*	*%*
Elaborate, cannot	I work mostly with languages that are not my native one, but even if I work with my native language, I always check what corpora say	5	26%
Starting point	Intuitions serve well when they provide initial hypotheses for researches	5	26%
Elaborate, can	Especially regarding low-resource languages, that provide few data.	3	16%
Depends	I believe intuition can stand alone as evidence only when we are dealing with clear cases, e.g. in English the article always precedes the noun.	3	16%
Evidence of X	Intuitions provide *a certain type* of evidence. They provide evidence of judgments about the question. But not about actual usage or processing, for example. It is metalinguistic knowledge.	2	10%
Focus on process	I do think syntactic intuitions are a good source of evidence, but we need to be careful on how these intuitions are gathered (number of speakers interviewed, way of doing it, etc)	1	5%
Total		19	100%

Significance (1):

Other, n = 0

Optional, n = 9:

Theme	*Example*	*No.*	*%*
When	All the time in grad school!	5	56%
Why	about the position of adverbs	3	33%
Uncategorised	[...] And, although I hardly ever incorporate reports of native speakers' intuitions in my written analyses, I do continue to solicit them when attemptin to understand particular phenomena. [...]	1	11%
Total		9	100%

Significance (2):

Optional, n = 10:

Theme	*Example*	*No.*	*%*
Elaborate, agree	Anyone who uses anything as evidence for anything should consider why their evidence can serve as evidence for their purposes.	5	50%
Follows from theoretical orientation	There's no point -- it depends on your presuppositions about the nature of grammar. If these are set out clearly, your position wrt the use of intuitions follows.	2	20%
Uncategorised	If they have no intuitions, why would they want to theorize?	2	20%
Starting point	Or perhaps, use these intuitions as hypotheses, which they then test using appropriate empirical data.	1	10%
Total		10	100%

Any other comments:

Optional, n = 16:

Theme	*Example*	*No.*	*%*
Comment on questionnaire	An interesting survey!	9	56%
Preferred data	I prefer corpus data (which entail a lot of methodological and theoretical problems as well). Syntactic intuitions is sometimes a useful complement.	4	25%
Central point	I reject the importance of being a 'native speaker'.	3	19%
Total		16	100%

Appendix D

Association analyses

Results of the association analyses performed for this dissertation are presented on the following pages. Only analyses with significant results are displayed. Analyses regarding participants' areas of specialisation were only carried out when more than 8 participants in the relevant group had chosen the particular specialisation (see an overview of participants' specialisations in section 5.5).

The method of analysis is described in section 6.3.1. In the tables, standardised residuals are presented below frequency counts and percentages. An alpha-level of .05 was used for all statistical tests presented in this appendix.

The order of presentation follows that of section 6.3. First, the results related to participants' theoretical orientation are presented (analyses were done both within the group of generative participants and within the group of all participants). Then results related to participants' areas of specialisation within linguistics are presented (analyses were done both for the large generative group, the small generative group, the non-generative group, and the group of all participants). Finally, results related to participants' methodological stance (on the experimentalist-traditionalist scale, see section 6.3.1 for an introduction) are presented (analyses were done both for the large generative group, the small generative group, the non-generative group, and the group of all participants). For a description of the relevant groups, see section 5.5.

D.1 Association analyses: theoretical orientation

The origin of syntactic intuitions:

Gen vs. non-gen: $p < .001$, $V = .43$

	Gen	*Non-gen*
Reflections	17 (23%) -4.93	40 (66%) 4.93
Competence	40 (55%) 3.74	14 (23%) -3.74
Other	16 (22%) 1.60	7 (11%) -1.60
Total	73 (100%)	61 (100%)

Acceptability and grammaticality:

No significant results.

The role of mental grammar:

Gen vs. non-gen: $p = .001$, $V = .34$

	Gen	*Non-gen*
Poor description	13 (18%) -3.51	28 (46%) 3.51
Good way to talk	44 (60%) 1.47	29 (48%) -1.47
Other	8 (11%) 1.27	3 (5%) -1.27
Actual process	8 (11%) 2.15	1 (2%) -2.15
Total	73 (100%)	61 (100%)

Fallibility of intuitions:[1]
Small gen vs. mixed: p = .04, V = 0.32

	Small gen	*Mixed*
Agree	15 (27%) 0.54	3 (20%) -0.54
Neither	6 (11%) -2.69	6 (40%) 2.69
Disagree	35 (62.5%) 1.57	6 (40%) -1.57
Total	56 (100%)	15 (100%)

Gen vs. non-gen: p < .001, V = .42

	Gen	*Non-gen*
Agree	18 (25%) -2.65	29 (48%) 2.65
Neither	12 (17%) -2.51	22 (36%) 2.51
Disagree	41 (58%) 4.86	10 (16%) -4.86
Total	71 (100%)	61 (100%)

[1] Two participants asked in the comments to have their answers removed from the analysis of this question, hence the lower number of participants.

The subject matter of the study of grammar:

Gen vs. non-gen: $p < .001$, $V = .35$

	Gen	*Non-gen*
Patterns in behaviour	26 (36%) -3.83	42 (69%) 3.83
Capacity in mind	33 (45%) 3.56	10 (16%) -3.56
Other	14 (19%) 0.68	9 (15%) -0.68
Total	73 (100%)	61 (100%)

Structure rules and mental rules:

Gen vs. non-gen: $p < .001$, $V = .36$

	Gen	*Non-gen*
Good hypothesis	21 (29%) 3.00	5 (8%) -3.00
As if	44 (60%) 0.53	34 (56%) -0.53
Other	3 (4%) -0.99	5 (8%) 0.99
Cannot infer anything	5 (7%) -3.27	17 (28%) 3.27
Total	73 (100%)	61 (100%)

The form of the implementation of mental rules:

No significant results.

Experts and ordinary speakers:

No significant results.

Data and theoretical virtues:

Gen vs. non-gen: p = .032, V = .22

	Gen	*Non-gen*
Agree	54 (74%) -2.40	55 (90%) 2.40
Neither	14 (19%) 2.47	3 (5%) -2.47
Disagree	5 (7%) 0.47	3 (5%) -0.47
Total	73 (100%)	61 (100%)

Gradience in grammar:

No significant results.

Traditional methods:

Gen vs. non-gen: p < .001, V = .38

	Gen	*Non-gen*
Agree	48 (66%) 4.37	17 (28%) -4.37
Neither	7 (10%) -1.67	12 (20%) 1.67
Disagree	18 (25%) -3.31	32 (52%) 3.31
Total	73 (100%)	61 (100%)

Experimental methods:

No significant results.

Syntactic intuitions and other types of linguistic evidence:

Gen vs. non-gen: $p < .001$, $V = .47$

	Gen	*Non-gen*
Can stand alone	64 (88%) 5.36	27 (44%) -5.36
Cannot stand alone	8 (11%) -5.39	33 (54%) 5.39
Other	1 (1%) -0.13	1 (2%) 0.13
Total	73 (100%)	61 (100%)

Significance (1):

No significant results.

Significance (2):

No significant results.

D.2 Association analyses: specialisations

The origin of syntactic intuitions:

Syntax:

All participants: p = .001, V = .32

	Not syntax	*Syntax*
Reflections	34 (61%) 3.61	23 (29%) -3.61
Competence	17 (30%) -1.99	37 (47%) 1.99
Other	5 (9%) -2.14	18 (23%) 2.14
Total	56 (100%)	78 (100%)

Non-generative group: p = .041, V = .32

	Not syn	*Syntax*
Reflections	29 (64%) -0.31	11 (69%) 0.31
Competence	13 (29%) 1.85	1 (6%) -1.85
Other	3 (7%) -1.98	4 (25%) 1.98
Total	45 (100%)	16 (100%)

Sociolinguistics:

All participants: p = .013, V = .26

	Not socio	*Socio*
Reflections	44 (38%) -3.03	13 (76%) 3.03
Competence	51 (44%) 2.04	3 (18%) -2.04
Other	22 (19%) 1.32	1 (6%) -1.32
Total	117 (100%)	17 (100%)

Cognitive linguistics:

All participants: p = .002, V = .31

	Not cogn	*Cognitive*
Reflections	38 (35%) -3.51	19 (73%) 3.51
Competence	48 (44%) 1.99	6 (23%) -1.99
Other	22 (20%) 2.01	1 (4%) -2.01
Total	118 (100%)	26 (100%)

Historical linguistics:

All participants: p = .010, V = .25

	Not hist	*Historical*
Reflections	47 (41%) -1.20	10 (56%) 1.20
Competence	52 (45%) 2.71	2 (11%) -2.71
Other	17 (15%) -1.96	6 (33%) 1.96
Total	116 (100%)	18 (100%)

Acceptability and grammaticality:

No significant results.

The role of mental grammar:

Semantics:

All generative participants: p = .012, V = .39

	Not sem	*Semantics*
Poor description	3 (7%) -3.02	10 (34%) 3.02
Good way to talk	28 (64%) 0.72	16 (55%) -0.72
Other	7 (16%) 1.67	1 (3%) -1.67
Actual process	6 (14%) 0.90	2 (7%) -0.90
Total	44 (100%)	29 (100%)

Small generative group: p = .008, V = .45

	Not sem	*Semantics*
Poor description	1 (3%) -3.10	9 (35%) 3.10
Good way to talk	19 (61%) 0.57	14 (54%) -0.57
Other	6 (19%) 1.78	1 (4%) -1.78
Actual process	5 (16%) 0.97	2 (8%) -0.97
Total	31 (100%)	26 (100%)

Fallibility of intuitions:[2]

Syntax:

All participants: p = .020, V = .24

	Not syn	*Syntax*
Agree	24 (43%) 1.49	23 (30%) -1.49
Neither	18 (32%) 1.44	16 (21%) -1.44
Disagree	14 (25%) -2.76	37 (49%) 2.76
Total	56 (100%)	76 (100%)

[2] Two participants asked in the comments to have their answers removed from the analysis of this question, hence the lower number of participants.

Typology:

All participants: p = .012, V = .26

	Not typ	*Typology*
Agree	43 (36%) 0.17	4 (33%) -0.17
Neither	27 (22.5%) -2.71	7 (58%) 2.71
Disagree	50 (42%) 2.26	1 (8%) -2.26
Total	120 (100%)	12 (100%)

Corpus linguistics:

All participants: p = .019, V = .24

	Not corp	*Corpus*
Agree	29 (30%) -2.28	18 (51%) 2.28
Neither	24 (25%) -0.44	10 (29%) 0.44
Disagree	44 (45%) 2.64	7 (20%) -2.64
Total	97 (100%)	35 (100%)

Theoretical linguistics:

Small generative group: p = .041, V = .34

	Not theo	*Theoretical*
Agree	6 (19%) -1.40	9 (36%) 1.40
Neither	6 (19%) 2.33	0 (0%) -2.33
Disagree	19 (61%) -0.21	16 (64%) 0.21
Total	31 (100%)	25 (100%)

The subject matter of the study of grammar:

Syntax:

All participants: p = .001, V = .32

	Not syntax	*Syntax*
Patterns in behaviour	38 (68%) 3.36	30 (38%) -3.36
Capacity in mind	9 (16%) -3.37	34 (44%) 3.37
Other	9 (16%) -0.28	14 (18%) 0.28
Total	56 (100%)	78 (100%)

Corpus linguistics:

All participants: p = .004, V = .28

	Not corpus	*Corpus*
Patterns in behaviour	43 (43%) -2.85	25 (71%) 2.85
Capacity in mind	39 (39%) 3.05	4 (11%) -3.05
Other	17 (17%) 0.004	6 (17%) -0.004
Total	99 (100%)	35 (100%)

Psycholinguistics:

Non-generative group: p = .010, V = .37

	Not psycho	*Psycho*
Patterns in behaviour	35 (80%) 2.90	7 (41%) -2.90
Capacity in mind	5 (11%) -1.71	5 (29%) 1.71
Other	4 (9%) -2.01	5 (29%) 2.01
Total	44 (100%)	17 (100%)

Structure rules and mental rules:

Morphology:

All generative participants: p = .007, V = .42

	Not morpho	*Morpho*
Good hypothesis	15 (30%) 0.34	6 (26%) -0.34
As if	32 (64%) 0.96	12 (52%) -0.96
Other	3 (6%) 1.19	0 (0%) -1.19
Cannot infer anything	0 (0%) -3.42	5 (22%) 3.42
Total	50 (100%)	23 (100%)

Semantics:

Non-generative group: p = .046, V = .37

	Not sem	*Semantics*
Good hypothesis	4 (9%) 0.16	1 (7%) -0.16
As if	29 (62%) 1.72	5 (36%) -1.72
Other	5 (11%) 1.27	0 (0%) -1.27
Cannot infer anything	9 (19%) -2.78	8 (57%) 2.78
Total	47 (100%)	14 (100%)

The form of the implementation of mental rules:

Theoretical linguistics:

Small generative group: p = .022, V = .60

	Not theo	*Theoretical*
Represented	8 (73%) 2.25	2 (22%) -2.25
Not represented	2 (18%) -2.67	7 (78%) 2.67
Other	1 (9%) 0.93	0 (0%) -0.93
Total	11 (100%)	9 (100%)

Experts and ordinary speakers:

No significant results.

Data and theoretical virtues:

Semantics:

All participants: p = .046, V = .21

	Not sem	*Semantics*
Agree	79 (87%) 2.36	30 (70%) -2.36
Neither	9 (10%) -1.41	8 (19%) 1.41
Disagree	3 (3%) -1.90	5 (12%) 1.90
Total	91 (100%)	43 (100%)

Gradience in grammar:

Syntax:

All participants: p = .010, V = .26

	Not syn	*Syntax*
Other	4 (7%) -2.46	18 (23%) 2.46
Extra-grammatical	12 (21%) -1.05	23 (29%) 1.05
Degrees of gr	40 (71%) 2.77	37 (47%) -2.77
Total	56 (100%)	78 (100%)

Theoretical linguistics:

All participants: p = .014, V = .25

	Not theo	*Theoretical*
Other	16 (16%) -0.22	6 (18%) 0.22
Extra-grammatical	20 (20%) -2.77	15 (44%) 2.77
Degrees of gr	64 (64%) 2.63	13 (38%) -2.63
Total	100 (100%)	34 (100%)

Traditional methods:

Theoretical linguistics:

All participants: p = .014, V = .26

	Not theo	*Theoretical*
Agree	41 (41%) -2.98	24 (71%) 2.98
Neither	16 (16%) 1.04	3 (9%) -1.04
Disagree	43 (43%) 2.33	7 (21%) -2.33
Total	100 (100%)	34 (100%)

Pragmatics:

All generative linguists: p = .026, V = .36

	Not prag	*Pragmatics*
Agree	42 (69%) 1.26	6 (50%) -1.26
Neither	3 (5%) -3.06	4 (33%) 3.06
Disagree	16 (26%) 0.70	2 (17%) -0.70
Total	61 (100%)	12 (100%)

Small generative group: p = .029, V = .36

	Not prag	*Pragmatics*
Agree	33 (72%) 1.66	5 (45%) -1.66
Neither	3 (7%) -2.71	4 (36%) 2.71
Disagree	10 (22%) 0.26	2 (18%) -0.26
Total	46 (100%)	11 (100%)

Experimental methods:

Cognitive linguistics:

All participants: p = .045, V = .22

	Not cogn	*Cognitive*
Agree	68 (63%) -2.50	23 (88%) 2.50
Neither	24 (22%) 1.68	2 (8%) -1.68
Disagree	16 (15%) 1.51	1 (4%) -1.51
Total	108 (100%)	26 (100%)

Theoretical linguistics:

All participants: p = .034, V = .22

	Not theo	*Theoretical*
Agree	74 (74%) 2.59	17 (50%) -2.59
Neither	16 (16%) -1.71	10 (29%) 1.71
Disagree	10 (10%) -1.60	7 (21%) 1.60
Total	100 (100%)	34 (100%)

Syntactic intuitions and other types of linguistic evidence:

Syntax:

All participants: p = .004, V = .27

	Not syntax	*Syntax*
Can stand alone	31 (55%) -2.64	60 (77%) 2.64
Cannot stand alone	25 (45%) 2.99	16 (21%) -2.99
Other	0 (0%) -1.21	2 (3%) 1.21
Total	56 (100%)	78 (100%)

Sociolinguistics:

All participants: p = .008, V = .28

	Not socio	*Socioling*
Can stand alone	85 (73%) 3.08	6 (35%) -3.08
Cannot stand alone	30 (26%) -3.27	11 (65%) 3.27
Other	2 (2%) 0.54	0 (0%) -0.54
Total	117 (100%)	17 (100%)

Corpus linguistics:

All participants: p = .002, V = .28

	Not corpus	*Corpus*
Can stand alone	75 (76%) 3.27	16 (46%) -3.27
Cannot stand alone	23 (23%) -3.11	18 (51%) 3.11
Other	1 (1%) -0.77	1 (3%) 0.77
Total	99 (100%)	35 (100%)

Significance (1):

No significant results.

Significance (2):

Pragmatics:

All generative participants: p = .041, V = .33

	Not prag	*Pragmatics*
Agree	56 (92%) 1.70	9 (75%) -1.70
Neither	3 (5%) 0.78	0 -0.78
Disagree	2 (3%) -2.72	3 (25%) 2.72
Total	61 (100%)	12 (100%)

Small generative group: p = .038, V = .39

	Not prag	*Pragmatics*
Agree	43 (93%) 2.01	8 (73%) -2.01
Neither	2 (4%) 0.70	0 -0.70
Disagree	1 (2%) -2.93	3 (27%) 2.93
Total	46 (100%)	11 (100%)

D.3 Association analyses: methodological stance[3]

The origin of syntactic intuitions:

All participants: p = .007, V = .27

	-2	*-1*	*0*	*1*	*2*
Reflection	22 (56%) 2.08	11 (48%) 0.56	19 (45%) 0.43	2 (11%) -2.90	3 (25%) -1.29
Competence	15 (38%) -0.28	8 (35%) -0.59	12 (29%) -1.87	13 (72%) 2.97	6 (50%) 0.72
Other	2 (5%) -2.37	4 (17%) 0.03	11 (26%) 1.87	3 (17%) -0.06	3 (25%) 0.75
Total	39 (100%)	23 (100%)	42 (100%)	18 (100%)	12 (100%)

Acceptability and grammaticality:

No significant results.

The role of mental grammar:

No significant results.

Fallibility of intuitions:

No significant results.

The subject matter of the study of grammar:

No significant results.

[3] The scores at the top of the tables are explained in section 6.3.1.3. Negative scores indicate an experimentalist stance on the methodology issue, and positive scores indicate a traditionalist stance.

Structure rules and mental rules:

All participants: p = .016, V = .24

	-2	*-1*	*0*	*1*	*2*
Good hypothesis	3 (8%) -2.20	1 (4%) -2.01	11 (26%) 1.34	8 (44%) 2.89	3 (25%) 0.51
As if	25 (64%) 0.89	13 (57%) -0.18	24 (57%) -0.17	9 (50%) -0.76	7 (58%) 0.01
Other	3 (8%) 0.54	1 (4%) -0.36	3 (7%) 0.39	1 (6%) -0.08	0 (0%) -0.91
Cannot infer anything	8 (21%) 0.82	8 (35%) 2.61	4 (10%) -1.46	0 (0%) -2.02	2 (17%) 0.02
Total	39 (100%)	23 (100%)	42 (100%)	18 (100%)	12 (100%)

The form of the implementation of mental rules:

No significant results.

Experts and ordinary speakers:

All participants: p = .002, V = .27

	-2	*-1*	*0*	*1*	*2*
Better evidence	3 (8%) -2.30	3 (13%) -0.93	8 (19%) -0.21	7 (39%) 2.13	6 (50%) 2.70
Equally good	11 (28%) -1.29	8 (35%) -0.20	18 (43%) 1.02	8 (44%) 0.75	4 (33%) -0.24
Worse evidence	21 (54%) 3.18	10 (43%) 1.10	12 (29%) -0.83	0 (0%) -3.24	2 (17%) -1.30
Other	4 (10%) 0.14	2 (9%) -0.18	4 (10%) -0.05	3 (17%) 1.07	0 (0%) -1.19
Total	39 (100%)	23 (100%)	42 (100%)	18 (100%)	12 (100%)

All generative participants: p < .001, V = .37

	-2	*-1*	*0*	*1*	*2*
Better evidence	0 (0%) -2.49	0 (0%) -1.28	5 (19%) -0.74	7 (44%) 2.19	5 (56%) 2.45
Equally good	4 (25%) -1.24	2 (40%) 0.08	12 (44%) 0.82	7 (44%) 0.50	3 (33%) -0.33
Worse evidence	10 (62.5%) 3.56	3 (60%) 1.69	6 (22%) -0.76	0 (0%) -2.78	1 (11%) -1.17
Other	2 (12.5%) 0.22	0 (0%) -0.81	4 (15%) 0.81	2 (12.5%) 0.22	0 (0%) -1.12
Total	16 (100%)	5 (100%)	27 (100%)	16 (100%)	9 (100%)

Data and theoretical virtues:

All participants: p < .001, V = .31

	-2	*-1*	*0*	*1*	*2*
Agree	37 (95%) 2.58	22 (96%) 1.94	34 (81%) -0.08	10 (56%) -3.02	6 (50%) -2.92
Neither	1 (3%) -2.26	0 (0%) -2.01	7 (17%) 0.94	5 (28%) 2.07	4 (33%) 2.25
Disagree	1 (3%) -1.07	1 (4%) -0.36	1 (2%) -1.18	3 (17%) 2.06	2 (17%) 1.64
Total	39 (100%)	23 (100%)	42 (100%)	18 (100%)	12 (100%)

All generative participants: p = .006, V = .34

	-2	*-1*	*0*	*1*	*2*
Agree	16 (100%) 2.69	4 (80%) 0.32	21 (78%) 0.57	8 (50%) -2.47	5 (56%) -1.34
Neither	0 (0%) -2.21	0 (0%) -1.13	6 (22%) 0.51	5 (31%) 1.39	3 (33%) 1.15
Disagree	0 (0%) -1.23	1 (20%) 1.21	0 (0%) -1.77	3 (19%) 2.13	1 (11%) 0.54
Total	16 (100%)	5 (100%)	27 (100%)	16 (100%)	9 (100%)

Gradience in grammar:

No significant results.

Syntactic intuitions and other types of linguistic evidence:

All participants: p < .001, V = .31

	-2	*-1*	*0*	*1*	*2*
Can	20 (51%) -2.64	12 (52%) -1.78	30 (71%) 0.59	17 (94%) 2.59	12 (100%) 2.50
Cannot	19 (49%) 2.92	10 (43%) 1.47	12 (29%) -0.34	0 (0%) -3.03	0 (0%) -2.41
Other	0 (0%) -0.91	1 (4%) 1.24	0 (0%) -0.96	1 (6%) 1.53	0 (0%) -0.45
Total	39 (100%)	23 (100%)	42 (100%)	18 (100%)	12 (100%)

Significance (1):

No significant results.

Significance (2):

Non-generative group: p = .049, V = .35

	-2	*-1*	*0*	*1*	*2*
Agree	23 (100%) 2.19	15 (83%) -0.82	12 (80%) -1.19	2 (100%) 0.52	2 (67%) -1.22
Neither	0 (0%) -1.61	3 (17%) 2.06	1 (7%) 0.02	0 (0%) -0.38	0 (0%) -0.47
Disagree	0 (0%) -1.38	0 (0%) -1.15	2 (13%) 1.74	0 (0%) -0.33	1 (33%) 2.33
Total	23 (100%)	18 (100%)	15 (100%)	2 (100%)	3 (100%)