ZS ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45 (1)

https://doi.org/10.18148/zs/2026-i1

Table of contents

Articles

Book review

Published by: publikon / dgfs (Linguistic Society of Germany)

Volume 44 (2) Spring Issue

Publisher

PubliKon
Deutsche Gesellschaft für Sprachwissenschaft – Linguistic Society of Germany

Co-Editors-in-Chief

Editors

Sarah Schimke
Department I - Germanistik, Komparatistik, Nordistik, Deutsch als Fremdsprache
Institut für Deutsche Philologie
Ludwig-Maximilians-Universität München
Schellingstraße 3 RG
80799 München

Melanie Weirich
Sprechwissenschaft & Phonetik
Friedrich-Schiller-Universität Jena
07737 Jena

Technical support

Miriam Butt
Technical Editor (OJS)
Department of Linguistics – Universität Konstanz
Universitätsstraße 10
78464 Konstanz, Germany

Sebastian Danisch
Publikon
https://www.kim.uni-konstanz.de/publikon/
78457 Konstanz, Germany

Jana & Matthias Müller
Copy editing

SiliconChips Services
Copy editing & production

Scientific advisory board

ZS ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 1

https://doi.org/10.18148/zs/2026-2001

Sophia Voigtmann, University of Bamberg, Germany, sophia.voigtmann@uni-bamberg.de

Processing German past tenses

The impact of sentence onset and dialect

Abstract: This study investigates the impact of the Uniform Information Density principle on the use and processing of German past tenses, focusing on the perfect and the preterit. It examines whether sentence-initial temporal versus tense-neutral adverbs affect acceptability and processing speed, and how regional variation – particularly the decline of the preterit in some dialects – modulates this effect. The findings from an acceptability judgement study and a self-paced reading study indicate that temporal adverbs ease the processing of preterit constructions, especially in dialect regions where the preterit is less common. It thus provides evidence for the influence of dialects on the processing of Standard German and supports the Uniform Information Density-based notion of optimal information distribution.

Keywords: Past tense, German, Uniform Information Density, dialects, rating experiment, self-paced reading experiment

Cite as: Sophia Voigtmann. 2026. Processing German past tenses: The impact of sentence onset and dialect. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi.org/10.18148/zs/2026-2001
Received: 16 July 2025; **accepted:** 02 January 2026; **published:** 08 June 2026
Published by: publikon / dgfs (Linguistic Society of Germany)

1 Introduction

Ever since the papers by Fenk & Fenk (1980), the concept of distributing information evenly across an utterance to facilitate successful communication has been a prominent topic in (psycho-)linguistics. This idea has gained further prominence in recent years for phonetic variation (Aylett & Turk 2004) and syntactic variation, particularly the omission or placement of optional words (Levy & Jaeger 2007).

The idea of a Uniform Information Density (UID) (Levy & Jaeger 2007) has been applied to various cases both in English and German (see Juzek 2024 for an overview). However, to my knowledge, the relationship between UID and the selection and usage of past tenses, especially in German, has not been extensively explored.

The past tenses locate an event at a time before the actual speech time (Reichenbach 1947), which is called 'PAST' throughout this paper. In German, the 'Perfekt' (present perfect) and the 'Präteritum' (simple past or preterit tense) can be used accordingly.[1] While there are semantic differences between these tenses (e.g. Musan 2002; Welke 2005), they are somewhat interchangeable (Hennig 2000: 26—29, among others). This interchangeability is particularly evident in German dialects because several of them have lost their preterit forms in their verbal paradigm (Fischer 2018: 1, 29). Consequently, the preterit is no longer used in these regions and is fully replaced by the perfect tense. Since dialects and the standard variant influence each other, it can be inferred that the interchangeability of the two tenses is also present in the standard German variant (Fischer 2018: 131—136; Hennig 2000: 29—31).

Moreover, past tenses are frequently accompanied by temporal adverbials, which – like the tense itself – convey past time reference. These adverbials often appear as frame- or scene-setters at the beginning of the sentence (Speyer 2008; 2009; 2010), positioning the temporal information early in the utterance. Building on this observation – and assuming that the preterit and perfect are fully interchangeable in isolated, contextless sentences – this paper examines how sentence-initial adverbials influence past-tense processing within the framework of the UID principle. Example (1) illustrates the idea.

(1) a. *Gestern hat die Angestellte die Blusen gebügelt.*
Yesterday has the employee the blouses ironed.
'Yesterday, the employee has ironed the blouses.'

[1] The Plusquamperfekt 'plusquamperfect' also locates an event in the past though it mostly fulfills the task of locating an event in the past before another event or reference time.

b. *Gestern bügelte die Angestellte die Blusen.*
Yesterday ironed the employee the blouses.
'Yesterday, the employee ironed the blouses.'

c. *Vielleicht hat die Angestellte die Blusen gebügelt.*
Perhaps has the employee the blouses ironed.
'Perhaps, the employee has ironed the blouses.'

d. *Vielleicht bügelte die Angestellte die Blusen.*
Perhaps ironed the employee the blouses.
'Perhaps, the employee ironed the blouses.'

The sentences in (1) should differ in terms of UID. In (1a) and (1b), the PAST information is spread across two units, the temporal adverb *gestern* 'yesterday' and the grammatical tense on the verb(s). This should facilitate processing of these clauses by spreading the PAST information across two words, especially in the preterit condition in (1b). In finite preterit verbs, lexical, valency and congruence information are clustered. This could result in an unwanted information peak when the PAST information is also only located in this single word (1d). However, in (1b), the PAST information is spread across the verb *and* the adverb, reducing the information load on the finite verb and, thus, smoothing out the information profile.

This paper aims to provide evidence for the influence of UID and tense, and seeks to answer the following questions: Are the perfect and preterit tenses perceived and processed differently depending on the sentence onset – with the preterit tense being rated higher and read faster when a temporal adverb occurs at the sentence onset? Can these differences be linked to UID? Are there differences depending on the dialect region of a participant due to the preterit loss there?

Thus, I have conducted a rating experiment and a self-paced reading study in a context-free setting to test the following more specific hypotheses using an acceptability judgement task and a self-paced reading experiment.

The remainder of this paper will be structured as follows: Section 2 focuses on the theoretical background about information theory (2.1), and the uniform information density (2.2). Section 3 presents differences between perfect and preterit tense (3.1), the preterit loss (3.2), and the adverbials at the sentence onset (3.3). Section 4 specifies the hypotheses, my expectations, and the processing claim about tense and sentence onset, and the overarching experimental setups. Section 5 presents the experiments and their results, which are discussed in Section 6. A summary in Section 7 closes the paper.

2 Surprisal and Uniform Information Density

2.1 Surprisal

In Shannon's (1948) fundamental, mathematical work on information theory, communication is described as a transfer of information. Communication is considered successful when no information is lost, even in challenging conditions such as noisy environments where individual words may not be comprehended. Nevertheless, communication as a whole does not necessarily break down with the misunderstanding of a single word. This is because words (or constituents) are not independent; rather, they influence each other's predictability, making certain elements like *car* in (2) more (2a) or less expected (2b). In more technical terms, this means that information transfer is successful when adding a certain amount of redundancy reduces the error rate of a signal to utilize the channel most optimally (Shannon 1948).

(2) a. I drive a car.
b. I buy a car.
c. I drive Susan's car.

There are fewer objects that can be *driven* (2a) compared to objects one can *buy* (2b). Thus, *car* is more expected after *drive* and carries less information, hence more redundancy, than after *buy*, where it is less expected.[2] Following the simplified account of Shannon's basic idea, hearers find reconstructing (2a) easier than (2b) due to the greater expectancy of *a car* after *drive*.

Accordingly, there is a negative correlation between the probability of a word and its informational content: The less likely a word is, the more information it conveys and vice versa. This information can be calculated and results in the so-called surprisal value (see Hale 2001 for the name and Shannon 1948: 379—389 for the definition): S(unit) = -log2 p(unit|context).

High surprisal values indicate more processing effort (e.g. Demberg & Keller 2008; Hale 2001) and depend on the amount of context, described as n-grams. Unigram models take only the frequency of a word into account (*car* in (2)). Bigram models include one previous word (*a* in (2a/b), *Susan's* in (2c)), trigram models two previous words (*drive/buy a* in (2a/b), thus the different expectancy would only be visible when one uses a trigram model) and so on.[3] Large and transformer-based language models can take even more context into consideration and use this information for prediction (Naveed et al. 2023). Though this study does not calculate surprisal

[2] Of course, this example is highly simplified and there are numerous contexts where the information what is driven is vitaly important.

[3] The general frequency of words (in a corpus) influences this value. Surprisal values are usually calculated using corpora that are divided into test and training data. Frequencies and likelihoods are gained in the training data and mapped on the test data (Manning and Schütze 2005 among others). Thus, the surprisal values often depend on the training data.

values to predict processing effort and will hence not elaborate further on the technical details of language models, the predictive power of certain elements is important. A less technical understanding of surprisal and prediction is given by Levy (2008).

He states that listeners and readers predict upcoming words based on the input they have comprehended so far (Levy 2008, among others). Even after only partial input, they keep various options about the sentence continuation in mind and assign different probabilities to them. Cognitive capacity is consumed when a probable continuation is ruled out, and the other possible clause continuations obtain a new (higher) probability assignment (Levy 2008: 1135—1136). In (2), it would be highly expected to find a determiner after both verbs. If this expectation was not met, for example, in (2c), more cognitive capacity would be consumed at *Susan's*, compared to the determiner *a*.

The examples mentioned above are on a lexical level. This notion of prediction and expectancy words, however, also applies on a more grammatical level. Thus, one would predict a direct object after the verbs in (2).

This also has implications for tense processing: once a temporal adverb such as yesterday is encountered, comprehenders are likely to rule out continuations that are incompatible with a past-time interpretation. As a result, a past tense verb becomes highly expected and should be processed more quickly and with less effort (Steinhauer & Ullman 2002).

2.2 The Uniform Information Density Principle

The previous section has shown that elements can differ in informativeness. These differences form the basis of the *Smooth Signal Redundancy Hypothesis* (Aylett & Turk 2004) and the UID Hypothesis (Levy & Jaeger 2007), which will be called *UID principle* throughout the paper. When one does not only consider single surprisal values but those of a whole sentence, the different values must be viewed in relation to one another.

Highly informative, i.e., very unexpected, content creates peaks in the information profile. Very uninformative, that is, highly expected, content results in a trough (Figure 1). This change from very high to very low surprisal values is considered very hard to process and should be avoided.

To avoid information profiles as in Figure 1, speakers have various techniques to smooth out the profile, for example, by distributing very informative content over a longer signal or more than one linguistic unit. Thus, information is transmitted "at a constant rate", which helps speakers to "ensure that their message is being communicated as rapidly as possible" (Collins 2014: 656).

Levy and Jaeger (2007) examine this issue in the context of English object relative clauses, where the relative pronoun can either be omitted or included. When omitted, the first word

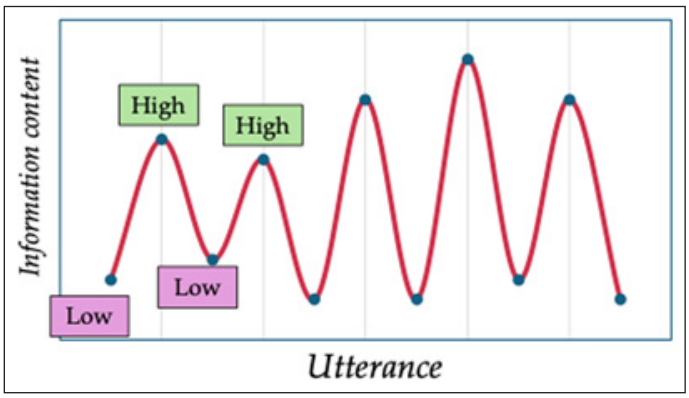

Figure 1: Schematic description of the change in low and high surprisal values, which results in an uneven information profile (Graphic taken from: Wall & Voigtmann 2025)

of the clause must both introduce the clause and carry its own lexical meaning. This dual function becomes problematic when that word is highly informative or when the relative clause is unexpected. Not using a relative pronoun might result in a change of low and very high information content, as indicated by Figure 1. Including the relativizer anchors the clause type explicitly, reducing the informational burden on the following word and making both it and the clause as a whole easier to process because the information profile is less "spiky" (Figure 2). A smooth information profile, as in Figure 2, is said to be preferable for processing. How this relates to the question of tense processing is shown in Section 4.1.

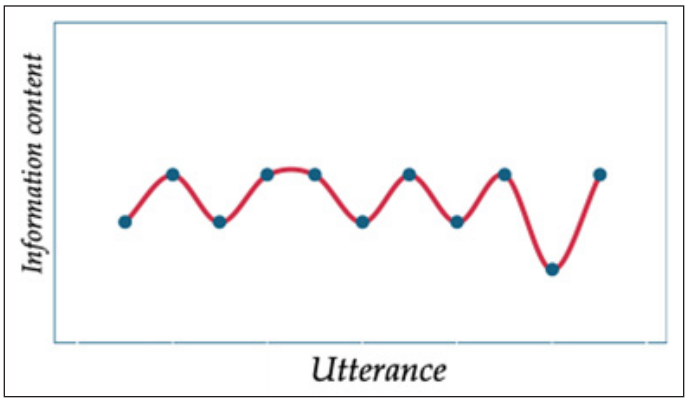

Figure 2: Schematic description of change in low and high surprisal values that results in an even information profile (Graphic taken from: Wall & Voigtmann 2025)

The UID principle has been successfully applied in various studies, e.g. on topic drop (Schäfer 2021, 2025) and fragments (Lemke 2021), but also on word order variation (Collins 2014; Ortmann et al. 2024; Wallenberg et al. 2021). However, there have also been mixed and negative results, as summarized, for example, by Juzek (2024).

3 Perfect and preterit tense in German

3.1 Perfect and preterit tense

Tense – especially past tense – is deictic, meaning it is interpreted relative to the time of speaking (Reichenbach 1947). As mentioned in the introduction, both past tenses considered here locate an event before the speech time (PAST).

This PAST location unifies preterit and perfect tense in modern German. However, when the reference time is also included, as demonstrated to be particularly useful to describe English tense by Rothstein (2006: 171—172), the perfect and the preterit differ. In preterit, the event time coincides with the reference time, making it suitable for referring to singular or repetitive past events and providing a simple context reference (Wöllstein & Bibliographisches Institut 2022: 214). The latter can be made explicit in examples like (3):

(3) We arrived in Oxford, which was by the River Thames.

At the 'speech time' of writing this article, *Oxford* is still located by the river *Thames*. Due to the location of reference time parallel to the event time, sentences like (3) are not ungrammatical and do not imply that the information of the context is no longer true.

Welke (2005: 299) emphasizes that the preterit is primarily used to indicate the temporal relevance of past events. However, he acknowledges its potential to express incompleteness and the progression of an event, contrasting it with the perfect tense.

In perfect tense, speech and reference time are parallel (Reichenbach 1947; Rothstein 2006: 25—27; Wöllstein & Bibliographisches Institut 2022: 210—212). This implies four interpretations as identified by Fischer (2018): futuristic, universal, perfective, and imperfective. Unlike the "präteritale" 'imperfective' interpretation of the perfect tense, which is essentially synonymous with the preterit usage (Fischer 2018: 133), the perfective interpretation signifies the temporal relevance of the past event for the present.[4]

The perfect seemingly carries a more intricate semantic interpretation than the preterit. It has been frequently proposed that the perfect tense yields a compositional interpretation; a present reading is visible in the auxiliary while a past or resultative meaning is attributed to the participle. This interpretation aligns with Musan's (2002) notion of the resultative state of the perfect tense and is further explored by Rothstein (2006: 130) to explain the connection between the perfect tense and events that still retain significance in the present. However, the possibility of the resultative state as an interpretation often depends on the verb's aspectuality,

[4] The other two categories are subordinate and will not be considered here.

as suggested by Welke (2005: 231—235). Aspect, not being a grammatical category in German (Vater 1994), but rather encoded in the verb stem or affixes (Leiss 1992: 30—45), is, thus, not coded via tense.

Welke (2005) expands upon Musan's (2002) interpretation of the perfect tense, but follows her on the relevance of the perfect event for the present speech time and on the notion of universality. Building on the latter, he adds a more external perspective on events to describe the totality of an event and to express completeness (Welke 2005: 299).

A different perspective is presented by Löbner (2002: 271), who argues for an ambiguity between "a non-past perfect and a past non-perfect". He bases his argumentation on several factors. Firstly, the perfect tense can occur in *als*-sentences ('when'), which are "incompatible with present or future time reference" (Löbner 2002: 267), and in *wenn*-sentences ('when'), that only have a "future perfect reading of the [perfect tense]". Additionally, perfect tense in *bevor*-clauses ('before') are also compatible with both readings. Perfect tense can also occur in *nach-dem*-clauses ('after'), but without a "present time reference" (Löbner 2002: 268). Secondly, he us es the compatibility of the perfect tense with the particles *schon* ('already') and *noch* ('still') (Löber 2002: 269–270).[5] However, independently of the occurrence of particles in a sentence, even the perfective aspect reading is situated completely in the past if a matching time referent like an adverb is given in the context (Löbner 2002: 260). Finally, Löbner (2002: 266–267) refers to the possibility of using the perfect tense in narratives, where the preterit is expected.

Preterit and perfect are generally associated with different contexts and registers. Preterit is linked to the written medium, including storytelling, reviews and private letters (Hennig 2000: 69, 72, 74), whereas the perfect is linked to the spoken medium, including dialouges in novels as well as audio and screenplays (Latzel 1977: 113).

Latzel (1977: 93—94) further suggests that the perfect tense is more commonly used with first and second person singular and in questions, which aligns with the oral medium's preference for the perfect tense and its spoken conception (Ágel & Hennig 2007; Koch & Oesterreicher 1985, 2007).

Perfect is also found in the written medium, for example, in law and instructional texts (Fischer 2018: 136), official letters (Hennig 2000: 69) or for structural purposes (Fischer 2018; Welke 2005). Hennig (2000: 65—69, 179—184) also examined talk show dialogues and found no preference for one tense over the other. However, these dialogues did reveal that certain verbs are more likely to be in the preterit tense than others. These verbs share an imperfective semantic.

[5] Note here, that neither noch nor schon were used in the items (see Section 4.2). Thus, I will not go into more detail about their different readings, but see Löbner (1989) for the distinction.

Complex constructions, such as the usage of auxiliaries or modal verbs, make preterit more likely (Fischer 2018: 381). This is attributed to the German sentence brackets, abbreviated as LSB (left sentence bracket) and RSB (right sentence bracket) here (Höhle 2018; Reis 1980; Wöllstein 2014). For instance, the usage of modal verbs fills both sentence brackets (4a), as does the perfect tense in general (4b).

(4) a. *Peter* $[wollte]_{\text{LSB}}$ *seine Hausaufgaben* $[machen]_{\text{RSB}}$.
Peter wanted his homework do
'Peter wanted to do his homework.'

b. *Peter* $[hat]_{\text{LSB}}$ *seine Hausaufgaben* $[gemacht]_{\text{RSB}}$.
Peter has his homework done
'Peter has done his homework.'

c. *Peter* $[hat]_{\text{LSB}}$ *seine Hausaufgaben* *[machen wollen]*$_{\text{RSB}}$.
Peter has his homework do want
'Peter has wanted to do his homework.'

When speakers use the preterit tense in these constructions (4a), they bypass complex constructions in the right sentence bracket, such as infinitive pro participio constructions (4c) (Askedal 1986a; 1986b; Bech 1955; Schallert 2014), while still maintaining the construction that distributes the syntactic predicate over both sentence brackets. A similar line of argument regarding the loss of the preterit is presented by Abraham (2004), as will be discussed in the following section.

The list of differences between preterit and perfect tense could be continued, especially with regard to their semantics (see, for example, Ogihara 2011). Nevertheless, it is possible to state a certain interchangeability of the two tenses, especially in the context of this study. Firstly, both refer to events that occurred before the speech time, excluding the subordinate and infrequent perfect functions. Secondly, they overlap to some extent and can lose their dividing effects in certain situations (Welke 2005: 297). Thirdly, there is no clear preference for perfect or preterit in contextless one-sentence-utterances (Hennig 2000: 29—31).

3.2 Preterit loss in the German Dialects

The standardized national variant of German is a rather modern phenomenon. Until the 19th century, dialects dominated everyday life and were even used in official communication (von Polenz 2013: 212—213, among others). They were gradually superseded by the standard variants. Even today, dialects persist in everyday communication and differ significantly from the standard variant.

The most prominent and well-documented regional differences in German are found at the phonetic and lexical levels (Niebaum & Macha 2014: 115–122; Wenker 1888; Wiesinger 1983). Syntactic variation also exists, and particularly relevant to this study is the loss of the preterit tense 'Präteritumschwund' in several, primarily southern, regions of Germany (Fischer 2018; 2021). In these areas, the preterit is no longer actively used or formed and has been largely replaced by the perfect tense. This is possible because, since its emergence in Old High German (750–1050 CE), the perfect tense has been able to take on more and more functions that were previously attributed to the past tense (Fischer 2018: 395; 2021: 351—353; Sapp 2009, among others). Thus, the decline of the preterit and the expansion of the perfect tense are closely linked (Fischer 2021: 351—353)

Based on dialect grammars, Fischer (2018) shows that in northern dialect areas, more preterit forms are documented while southern German regions only obtain preterit forms occasionally, e.g. for *to be* or modal verbs. The documentation of preterit forms corresponds to the usage frequency of the tense; in northern regions, preterit is used more often than in southern regions (Fischer 2021: 336). This North-South divide is also evident in Kasper and Pheiff's (2023: 43–47) usage-based study. Though preterit forms can be found all over Germany, they are more frequent in the northern German regions compared to the southern German regions.

There is, however, a transition zone from Rhine Franconian to Moselle Franconian, to Ripuarian, to northern Hessian in the West and to East Franconian in the east (Fischer 2021: 333), where more preterit forms are kept. This shows that the isoglosses of preterit usage are not identical to the dialect isoglosses themselves; only a north-south divide can be observed (Fischer 2021: 335).[6]

Besides the areal distribution of preterit loss or maintenance, certain verbs keep preterit better than others. This phenomenon applies to frequently used verbs, syntactically intricate verbs that construct complex predicates, morphologically irregular verbs, and verbs related to states and activities (Henriksson 2006: 45). Consequently, mostly strong and irregular verbs, as well as preterit present verbs, tend to retain their preterit form. This observation corresponds to the token frequency of these verbs in a spoken German corpus comprising both dialectal varieties and forms closely approximating Standard German, although the verbs in question are not necessarily among the most frequently occurring overall (Fischer 2021). Fischer (2021: 343) argues that the irregular preterit forms are more lexicalized than the regular weak verb forms, more economical in articulation and are more distinguishable from their present form due to the stem vowel change. These results are also confirmed by Kasper and Pheiff (2023: 40–41); the strong and frequently used verb *kommen* ('come') is chosen and translated more often as preterit by speakers of dialects, regional dialects, and standard German than the weakly inflected and less common word *wohnen* ('live').

[6] However, there is a geographic overlap – to some degree – between preterit maintenance regions and areas that have kept the apocope in the dialects (Fischer 2018: 331—332).

The preterit preference for verb types like modals, that build complex predicates even with preterit and present tense,[7] is linked to psycholinguistic and processing explanations, though not tested with this methodology (Abraham 2004).

Abraham (2004: 243—244) argues that in Upper German, where the preterit has been entirely lost, the use of the perfect tense and the "tun-periphrasis" consistently results in the verb being distributed across both sentence brackets. He claims that this structural pattern facilitates processing as it spares recipients from having to adjust to shifting verb positions in verb-second clauses, among other factors not addressed further here. Crucially, Abraham (2005: 119–122) later also suggests that distributing the predicate across both brackets enables earlier identification of the subject via the finite verb in the left bracket, which agrees with it. However, this argument ignores the fact that a finite full verb is also congruent with the subject and thus performs at least a similar identification function. Moreover, the advantage described above should only come into play when differentiating the person of the subject (e.g. first person vs. third person), because fewer syncretisms occur with certain verbs (e.g. *to be*), but also not with all verbs in not all tenses. The predictive character of auxiliary verbs in the left clause bracket is also very limited overall, as Voigtmann (accepted) points out with reference to Levy (2008). If no valency information appears in the left clause bracket, there are more possibilities for filling the middle field, but these are more difficult to predict, which in turn implies processing difficulties. Consequently, it is not surprising that Sapp (2009) does not find empirical evidence for this claim in historical data, and Drinka (2004) also provides evidence against it. Fischer (2021: 343), nonetheless, builds upon Abraham's (2004; 2005) ideas to explain why modals and *sein* ('to be') maintain their preterit usage effectively and suggests that using them in the preterit tense prevents the right sentence bracket from becoming overly complex, as described above.

Despite obvious problems with Abrahams' (2004; 2005) ideas, I will also take the spread of the predicate across both sentence brackets into consideration here, as it has implications for UID, which will be explored in section 4.1. after discussing adverbials.

3.3 Adverbials at the sentence onset

Following the ideas already presented in the introduction, this section examines adverbials, which are among the most frequent elements found at the sentence onset (Speyer 2008; 2010).

Syntactically, adverbials are constituents that can be realized as noun phrases (NP), prepositional phrases, adjective phrases, clauses, or adverb phrases (e.g. van der Auwera 1998) that are used in this study and function as adjuncts (Frey 2003). Semantically, adverbials usually serve as

[7] These tenses are the only two synthetic verb forms in German if they are presented in active voice.

modifiers, providing contextual specification for the verbal or sentential reference (Maienborn & Schäfer 2011: 1391).

Maienborn & Schäfer (2011: 1393, highlighting by S.V.) classify the following adverbials: "*[p]redicational adverbials* assign a (gradable) property to the verbal or sentential referent they combine with. *Participant-oriented adverbials* introduce a new entity that takes part in the eventuality described by the verb. '*Functional adverbials*' is the cover term for the remaining adverbials."

Within predicational adverbials, sentence adverbials constitute a notable subclass. I focus on the further subclasses of epistemic and evaluative adverbials here (Maienborn & Schäfer 2011: 1396). Epistemic adverbials (e.g. *maybe*, *surely*, *probably*, *perhaps*) express the speaker's assessment of the truth value of the proposition (5a). Evaluative adverbials (e.g. *luckily*) convey the speaker's attitude toward the situation described in the sentence (5b). They "cannot appear in the scope of sentence negation [or] questions" (Specht 2023: 37) and differ in veridicality. According to Schäfer (2013: 39—42), the veridicality of epistemic adverbials – that is, their implication for the truth of the proposition – is scalar. Adverbials expressing higher certainty are considered more veridical than those indicating uncertainty (e.g. Ernst 2009). Evaluative adverbials, on the other hand, are always veridical.

(5) a. Maybe/ surely/ probably/ perhaps, Susan made the cake.
b. Luckily, Susan reads a lot.

Crucially for this study, sentence adverbials are frequently and unproblematically placed in the sentence-initial position and are not tied to a specific tense, as illustrated in (5). For this reason, they are classified here as tense-neutral (or simply neutral) adverbials.

Even more commonly found at the sentence onset are scene- or frame-setting adverbials (Speyer 2008: 285). Speyer (2008: 280) defines them "as an expression that names a crucial restriction on the situation (such as: the place, the time, etc. . .) in which the proposition is true", which aligns with the definition by Maienborn & Schäfer (2011: 1400): "[F]rame adverbials […] set a frame for the overall proposition."

The most relevant frame-setters here are temporal adverbials, which are often inherently linked to tense.[8] Interestingly, temporal adverbials appear more frequently at the sentence onset when they serve the restrictive, scene-setting function. This highlights their relevance to the current investigation. Unlike sentence adverbials, temporal adverbials are both non-veridical and tense-specific. When used, they enable both speaker and hearer to anticipate the requirement of a past tense verb. Ogihara (2006), among others, even assumes that only they, rather than tense affixes,

[8] Speyer (2008) also mentions the exitance of temporal adverbials that do not convey this meaning.

carry tense information either overtly as in (1a/b) or covertly. Even when one does not consider this formal framework, overt temporal adverbials distribute the PAST information between the temporal adverbial and the finite verb. This early signalling aligns with the principle of UID, whereby spreading information reduces processing difficulty and prevents information overload, thus facilitating efficient communication. In contrast, sentence adverbials do not serve this anticipatory or distributive function with respect to tense, which makes them good candidates to be contrasted with temporal adverbials.

Nonetheless, the following points should be critically noted: According to Maienborn & Schäfer (2011), sentence and temporal adverbials belong to two different classes, differ in their veridicality, are negated differently and must be accommodated differently in questions. Nevertheless, I have deliberately decided to use these two classes because they take scope over the elements below them (Frey 2003), occupying the highest positions in the sentence. It is not crucial here that the position of the sentence adverbials is even higher than that of the frame-setters because I do not combine the two adverbials in a sentence (see Specht (2023) for a processing account of combined adverbials). Secondly – and perhaps more importantly – both classes should be perceived as unmarked at the beginning of the sentence and should not introduce contrastivity. Contrastivity and the high likelihood of a focus reading prevent the usage of other frame-setters like local adverbs. The focus sensitivity of temporal adverbs has mostly faded (Eckardt & Speyer 2014: 517), and I do not assume a contrastive focus reading for sentence adverbs because the only likely set of alternatives (e.g. Rooth 1992) are other sentence adverbs. Consequently, the use of sentence adverbials allows possible effects to be attributed to PAST information rather than focus.

4 Experimental setup

4.1 Hypotheses and expectations

To answer the research question stated in Section 1, this paper investigates the following hypotheses:

(H1) In context-free settings, German sentences in the perfect or preterit tense are rated as differing in naturalness depending on whether they begin with a temporal or a sentence adverb.

(H1.1) Preterit tense sentences with a temporal adverb are rated as more natural than those with a neutral adverb.

(H2) The ratings differ depending on the dialect region of the participants.

(H3) A temporal adverb at the sentence onset facilitates processing of the finite verb, especially in the preterit tense and depending on the dialect region of the participants.

One of the most evident differences between the perfect and preterit tenses lies in their construction. The perfect tense is analytic, formed by combining an auxiliary verb and the past participle of the lexical verb. In German main clauses, the lexical verb is positioned in the right sentence bracket, almost at the end of the clause, while the auxiliary verb occupies the second position, in the left sentence bracket (e.g. Wöllstein 2014). On the other hand, the preterit tense is synthetic, indicating the tense through a suffix for weak verbs and/or stem vowel change for strong or irregular verbs (Wöllstein & Bibliographisches Institut 2022: 665).

In contrast to the perfect tense, the lexical, valency, modus, congruency, and tense information are encapsulated within a single word in the preterit tense and are processed in this single word. This concentration of information can lead to an unwanted information peak (see Figure 3), which may hinder processing according to the UID principle (Collins 2014; Levy & Jaeger 2007; Meister et al. 2021). However, there are several strategies to mitigate this issue. For instance, placing the subject in front of the finite verb addresses congruency information (Schäfer 2021), and the speech time reference issues can also be resolved earlier, as discussed in the introduction.

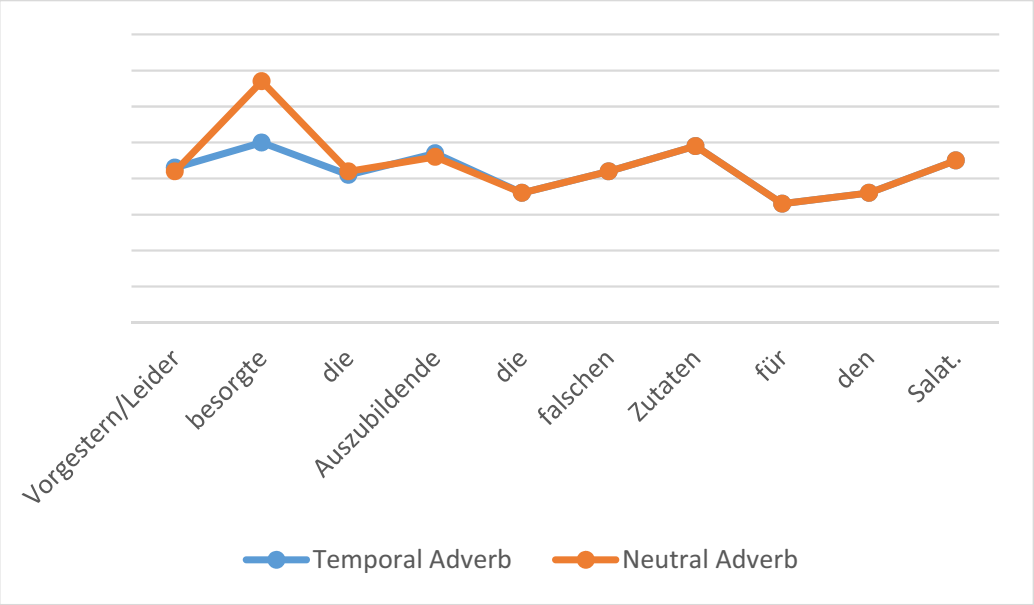

Figure 3: Schematic plot of the expected surprisal value differences in preterit sentences depending on the sentence onset (example 6b/c)

In general, the advantage of giving the PAST information early should apply to both the preterit and perfect tenses. However, the perfect tense is analytic, with the lexical verb positioned at the end of the clause, where locality effects can be observed and facilitate processing (Levy

2008). Consequently, there is less information on the finite verb, leading to a natural spread of information.[9] As a result, the positive processing effect of the early PAST information should be more pronounced for the preterit tense compared to the perfect tense.

There is one additional factor coming into play here: The preterit loss regions. It is well known that frequency (e.g. Díaz-Campos & Balasch 2023; Divjak 2019) as well as habituation to a phenomenon (Frank et al. 2016, among others) influences processing. It can also influence acceptability judgements as people might show a preference for what is more familiar to them. Hence, people from regions that have lost the preterit tense have less exposure to it compared to those who maintain it, since they read preterit but are less likely to hear it in Standard German or dialectal conversations. This reduced familiarity can lead to processing difficulties and differences in the acceptability judgement.

Accordingly, I propose that recipients from the preterit loss regions profit more from the information spread introduced by the temporal adverbial at the sentence onset with preterit case. For them, the missing familiarity with the preterit tense must be added to all information conveyed in the finite preterit verb. Consequently, the information should be perceived as even more significant for them.

It should also be mentioned that reading times are rarely measured that early in a clause. However, articles that instruct readers on self-paced reading experiments (e.g. Jegerski 2014; Mitchell 2018) make no statements about the position of the critical region. As there are effects visible, as will be shown in Section 5.2, this should not be a major concern.

4.2 Experiments

To test the hypotheses, two different experiments are used: an acceptability judgement task for H1 and H2 and a self-paced reading experiment for H3.

Acceptability judgement or simple rating studies are classically represented by questionnaires. Participants read sentences like those presented in Section 4.3 or in example (1) and judge the acceptability of sentences either on a scale or categorically. In this experiment, a 7-point Likert Scale with 7 being completely natural and 1 being completely unnatural is used. The experiment was conducted online using PCIbex (Zehr & Schwarz 2022). After an introduction explaining the task to judge the naturalness of the sentences, without giving information about the research question, participants saw single sentences and the Likert-Scale on their screen at home. Having assessed the sentences, they could continue to the next one at their own speed.

[9] Processing difficulties on the constituents in the middle field may, nevertheless, apply, but will not be considered here.

Acceptability judgement experiments like this are so-called offline experiments (Sichelschmidt and Carbone 2003: 116) and cannot make direct statements about processing at a specific point of the sentence, as the internal mechanisms governing the judgements stay hidden from the experiment leader; one can only observe the result (Sichelschmidt and Carbone 2003: 116).

In the context of the UID principle, it is expected to observe better judgements for sentences with a more uniform profile because the participants intuitively show a preference for informationally well-formed sentences (for another application of an acceptability judgement task in UID-related studies, see Schäfer 2025: 338–354 or Lemke 2021: 189–199).

The second experiment is a self-paced reading study. This belongs to the on-line group of experiments, more precisely a reaction time experiment (see e.g. Scharlau et al. 2003). In contrast to the acceptability judgement task, participants can assess processing (Scharlau et al. 2003: 190) because they measure how long it takes for a reaction to occur after a stimulus. Longer reaction times indicate processing difficulties (Scharlau et al. 2003: 191).

In the self-paced reading design used here, participants read sentences online on a computer at home. Again, the platform for the experiment was PCIbex (Zehr & Schwarz 2022).

These sentences are initially given only by lines. Each line represents a word and is exactly as long as the word (Figure 4). When participants press the space bar, a word appears in place of the line. If they press the space bar again, this word disappears, and the next one appears, which is called a linear and non-cumulative design (Jegerski 2014). Because the presentation is word-for-word, reaction time differences can also be measured for each word, allowing for a detailed determination of where difficulties arise.

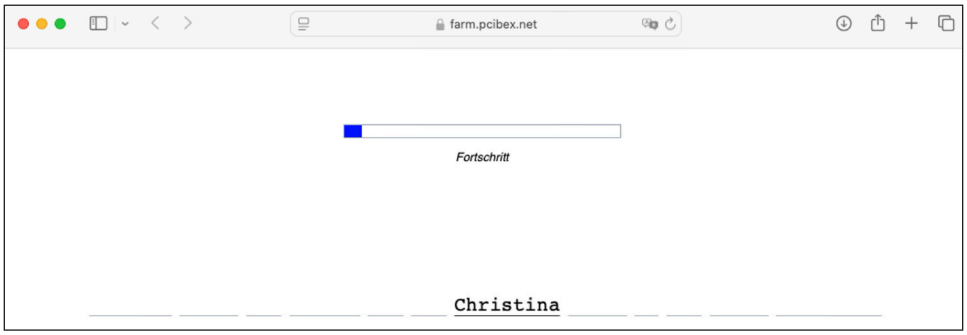

Figure 4: Screenshot of a sentence from the practice phase of the self-paced reading experiment

In an instructional part of the task, the participants were asked to read the sentences carefully and at their usual reading speed to understand the content. The latter was tested with polar comprehension questions after every sentence. These could be answered by pressing “j” for

Yes and "f" for No. Before the actual experiment, there was a practice phase of four items to familiarize the participants with the settings.

4.3 Items

The experimental design used in the two experiments has a 2 × 2 design[10] with 24 token sets. The two factors are tense (preterit vs. perfect) and the sentence onset (temporal adverb vs. tense-neutral adverb).

All items are simple main (verb-second) clauses. The subjects are in the third person singular since the first and second person are more frequent in perfect (Latzel 1977). To have a comparable spillover region across items behind the critical region of the finite verb (Mitchell 2018), the subjects are job descriptions with a definite article (e.g. *the physician*). Only transitive or intransitive weak verbs with one argument in the accusative or oblique case (PP) were used. If necessary, adjuncts have been added to make the sentences sound more natural and to prevent acceptability ratings from depending on clumsy sentences rather than the experimental conditions.

For the sentence onset condition, I included the temporal adverbs *gestern* 'yesterday', *letztens* 'recently', *neulich* 'the other day' and *vorgestern* 'the day before yesterday' (6c/d) and for the sentence adverbs *bestimmt* 'surely', *leider* 'unfortunately', *vielleicht* 'perhaps', and *wahrscheinlich* 'probably' (6a/b) were used.

(6) a. *Leider hat die Auszubildende die falschen Zutaten für den Salat besorgt.*
Unfortunately has the trainee the wrong ingredients for the salad got.
'Unfortunately, the trainee has got the wrong ingredients for the salad.'

b. *Leider besorgte die Auszubildende die falschen Zutaten für den Salat.*
Unfortunately got the trainee the wrong ingredients for the salad.
'Unfortunately, the trainee got the wrong ingredients for the salad.'

c. *Vorgestern hat die Auszubildende die falschen Zutaten für den Salat besorgt.*
the-day-before-yesterday has the trainee the wrong ingredients for the salad got.
'The day before yesterday, the trainee got the wrong ingredients for the salad.'

[10] This design classification refers to a multifactorial crossed design. Thus, both the effect of a single factor (tense or sentence onset) and the interaction between the factors (e.g. whether the influence of tense depends on a certain sentence onset) can be interpreted (see Sichelschmidt and Carbone 2003: 119–120).

d. *Vorgestern* *besorgte* *die* *Auszubildende* *die* *falschen* *Zutaten*
the-day-before-yesterday, got the trainee the wrong ingredients
für *den* *Salat.*
for the salad.
'The day before yesterday, the trainee got the wrong ingredients for the salad.'

As distractors for the experiments, each participant saw 74 sentences, including eight ungrammatical catch phrases and ten fillers. The fillers are from two other experiments about gapping and sluicing in German. They are in present tense and share no critical lexical material with my items.

4.4 Participant recruitment and allocation to the preterit loss region

The participants were recruited through Prolific (Prolific 2024) and paid about 4€ for the 20 minutes that each experiment took on average. The sum is based on the standard minimum wage (€12.50/h) in Germany. To keep the personal information to a minimum, participants were only asked to state whether their native language is German and name the *Landkreis* 'county' they have lived longest in.

The question about the county serves as an approximation to the preterit loss regions. In classical dialect syntactic research, which is strongly focused on basic dialects (Abraham & Leiss 2016; Fleischer et al. 2012; Niebaum & Macha 2014; Weiß 1998) and usually regionalizes rating items (Fleischer et al. 2012: 10—12), the county would not be a sufficient localization for the dialect, especially because administrative regions rarely correspond to isoglosses.[11] Additionally, the preterit loss isoglosses cannot be transferred one-to-one to the individual dialects in any case (Fischer 2018; 2021).

Of course, this type of data collection jeopardizes a controlled distribution across the two preterit loss regions. In fact, the participants were not evenly distributed across the experiments; the proportion of preterit maintenance region (PMR) participants was higher in both the rating and the self-paced reading experiment. However, unless the test subjects are specifically approached in the individual regions, which drastically extends the duration of the data collection, this problem cannot be solved, as no corresponding, census-based pre-screening outside the UK and the USA is available via Prolific.

A further argument for the approximation via the counties and the length of stay is that the Preterit Loss is not a phenomenon directly recognizable as dialectal for untrained individuals as it does not rely on a certain pronunciation or lexicon. This also speaks in favor of the transferability to

[11] Rheinfranconian can serve an example here. This dialect is spoken in parts of Saarland, parts of Rhineland-Palatinate and parts of Hesse and even in a small part of Bavaria (e.g. Wiesinger 1983: 847).

Standard German, which is the focus here. If a participant has spent most of their life in a Preterit Loss region (PLR), I assume it is very likely that they will adapt to the customs of this region regardless of whether they originally come from a Preterit Maintenance region. In concrete terms, this means that they should find preterit unusual in (oral) conversation and generally use it less frequently, i.e. only in written texts from genres typically associated with preterit like narratives.

Having received the county information, the county was located on a map of Germany and mapped to the preterit loss or maintenance regions in Fischer (2021: 349). Fischer's (2018; 2021) data is, however, based on dialect grammars from the 19th and early 20th centuries. Furthermore, she claims that the preterit loss might already have progressed into the transition zone. Within this transition zone, she describes a gradation depending on the verb type (Wöllstein & Bibliographisches Institut 2022: 357–359): In the northern part of the transition zone, preterit is lost only with weak and rare verbs. That means that preterit is kept with strong and modal verbs and, thus, technically with a larger group of verbs than in the southern region of the transition zone where preterit is also lost with modals and strong verbs. This is relevant because the predictions made here depend heavily on frequency and habituation. Participants in the southern transition zone, therefore, encounter the past tense less frequently than those in the northern transition zone.

Thus, I chose to divide the participants into two groups: Preterit maintenance region (PMR) north of Mainz (about 50° 0′ 0.0" N) and preterit loss regions (PLR) south of 50° 0′ 0.0" N. This is again only an approximation as the county borders are not exactly mappable to this geographic coordinate of latitude, but it should work well enough for the purpose of this study.[12]

5 Results

5.1 Experiment 1: rating study

5.1.1 Results

48 self-proclaimed native speakers of German, of which one was excluded due to a lack of county specification, (PMR: n = 30, PLR: n = 17) were asked to rate the grammatical acceptability of the sentences on a 7-point Likert Scale with 7 being completely natural and 1 being completely unnatural.

[12] As it was requested by a reviewer, I also ran the statistics, presented in the following section, with a subset of participants excluding those who have lived longest in the transition zone. For the rating experiment, that meant, however, the exclusion of 18% (n = 9) and for the self-paced reading experiment of 20% (n=20) of the participants. This reduces the reliability of the data which is one reason why I will present the data with all participants here. The other reason is that the results did not change significantly from those presented here.

To ensure transparency, however, the data and statistics on the reduced data set will still be made available (see Section 8).

The ratings were expectedly high (Table 1, Figure 5), indicating that all sentences are perceived as grammatical and rather natural. This was intended by the choice of onset. Also, the differences in the mean ratings are small. The highest ratings are found for items with temporal adverbs (6.55); the lowest ratings are given for sentences with a neutral onset and preterit in both dialect regions (6.05, 6.25).

Looking at the low ratings, one finds that the effect is driven by the ratings of items beginning with *vielleicht* 'perhaps'. They receive a mean rating below 6 (Table 2) in every condition, though they also belong to the group of "evaluating sentence adverbs" (Nübling 2016: 586). *Leider* 'unfortunately' was expected to be the odd-one-out in that group as it does not express a degree of uncertainty (Hoberg 1973) but comments on the speaker's attitude towards the proposition (Nübling 2016: 586). Possible explanations for the deviation of results are discussed in Section 5.1.2.

Table 1: Mean Ratings per condition and dialect region with confidence interval, experiment 1

Adverb	Tense	Region	Mean Rating	95% confidence interval	
Neutral	Perf	PLR	6.32	6.20	6.45
Temporal	Perf	PLR	6.30	6.19	6.42
Neutral	Pret	PLR	6.05	5.92	6.18
Temporal	Pret	PLR	6.39	6.27	6.49
Neutral	Perf	PMR	6.43	6.30	6.55
Temporal	Perf	PMR	6.55	6.44	6.67
Neutral	Pret	PMR	6.25	6.12	6.39
Temporal	Pret	PMR	6.55	6.43	6.66

Table 2: Ratings for *vielleicht* 'perhaps' in experiment 1

Adverb	Tense	Region	Mean Rating	95% confidence interval	
vielleicht 'perhaps'	Perf	PLR	5.82	5.70	5.95
	Pret	PLR	5.56	5.44	5.67
	Perf	PMR	5.98	5.85	6.11
	Pret	PMR	5.95	5.84	6.06

To test whether these differences are statistically significant, I used a cumulative link mixed model (Christensen 2023) in R (R Core Team 2024). The dependent variable was the rating result. The independent predictors were the sum-coded sentence onset (neutral −0.5, temporal 0.5), the

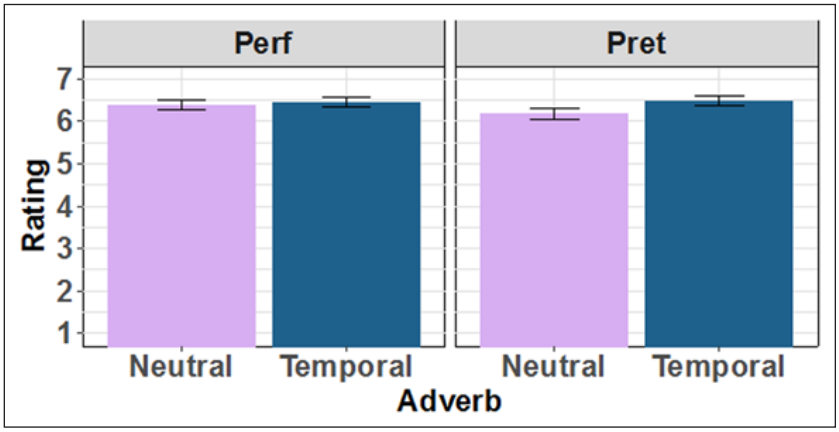

Figure 5: Mean Ratings, experiment 1. All plots are created with ggplot 2 (Wickham 2016)

sum-coded tense (perfect −0.5, preterit 0.5), the dialect region (PMR −0.5, PLR 0.5), the centered and scaled position of the items and the three-way interaction between onset, tense and dialect region. As random effects (Table 4), I included the random intercept of the participants, the nested random effects of onset and item to account for the variation caused by *vielleicht* and other adverbs. The results of the analysis are presented in Table 3.

Table 3: Results of the regression analysis, experiment 1. AIC: 2075.08

	Estimate	Std. Error	z-value	p-value	
Adverb	0.53	0.39	1.36	0.17	
Tense	−0.28	0.19	1.45	0.15	
Dialect region	0.46	0.45	1.03	0.30	
Position	−0.001	0.002	−0.58	0.56	
Adverb:Tense	0.77	0.38	2.00	0.05	*
Adverb:Dialect Region	0.099	0.29	0.34	0.73	
Tense:Dialect Region	0.07	0.29	0.25	0.81	
Adverb:Tense:Dialect Region	0.15	0.58	0.26	0.79	

The interaction between tense and the adverb at the sentence onset is significant ($z = 2.00$, $p < 0{,}05$, Figure 6). The rating of sentences differs depending on the tense *and* the sentence onset. The sentence onset does not significantly influence the rating of perfect sentences, but preterit sentences are more acceptable when they start with a temporal compared to a tense-neutral adverb.

Table 4: Random effects, experiment 1

Group	Variance	Std. Dev.
Item:Onset	0.379	0.6156
Participants	1.808	1.3447
Onset	0.225	0.4745

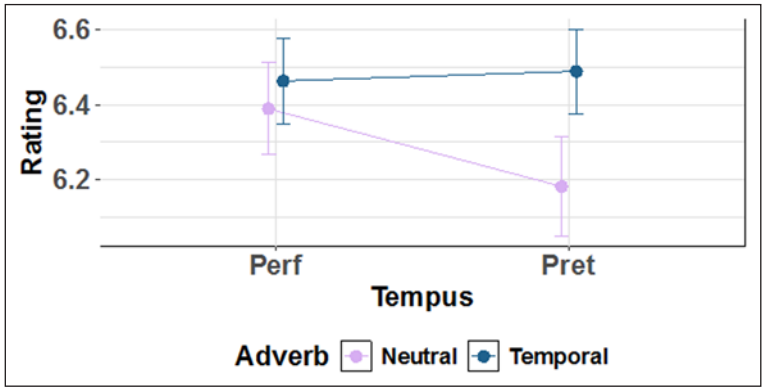

Figure 6: Interaction plot between tense and sentence onset. The y-axis displays the mean ratings

5.1.2 Discussion

The items in experiment 1 were generally very acceptable, as the item design was meant to be unobtrusive. As already mentioned, both sentence adverbs and temporal adverbs are among the most common words at the onset and are, thus, expected to be best suited to manipulate the tense information without confounding the results by adding another unwanted layer of interpretation.

However, this was apparently not as successful as anticipated, as the sentence adverbials showed an unexpected variation in their ratings. I originally anticipated *leider* 'unfortunately' to show the highest degree of variation since it belongs to the class of evaluating sentence adverbials (Nübling 2016: 586), whereas the other three adverbs express how likely the speaker considers an event to be. However, this was not the case as its mean ratings were quite close to those of all other adverbs, including the temporal adverbs.

However, it was *vielleicht* 'perhaps' that showed the greatest variability and the lowest acceptability. The items beginning with *vielleicht* express the highest degree of uncertainty about the clauses' propositions and evoke a contrastive focus reading for the subject (7).[13] Without

[13] Maienborn & Schäfer (2011: 1396) state "that maybe is […] not gradable but shares the general characteristics of the other predicationals used here." Why the gradability or lack of it should influence the ratings must remain an open question here.

context, these clauses might seem odder than the other items beginning with a sentence adverb, and the rating might reflect this rather than the intended conditions. However, in using the adverbs as random effects, the model can include this variability.

(7) *Vielleicht hat die Sekretärin den Tisch im Restaurant reserviert.*
perhaps has the secretary the table in restaurant booked
'Perhaps, the secretary has booked the table in the restaurant.'

The statistical analysis has shown a significant interaction between the sentence onset and the tense, which is in line with the predictions about the UID's influence. The temporal adverb spreads the PAST information across two words: the adverb itself and the predicate. This is especially helpful in the preterit sentences, where lexical, valency modus, congruency (since the subject is placed behind the finite verb, and first and third Person Singular showcase syncretism in the preterit verb), and tense information must be processed on one word if additional processing aids like temporal adverbs or adverbials are missing. In spreading at least the tense information, the unwanted information peak is avoided, which is generally preferred according to UID (Levy & Jaeger 2007).

However, the results must also be interpreted with a certain caution. Fischer (2018: 135) claims that the existence of a temporal adverb like *da*, *dann*, *damals* ('then') increases the likelihood of observing preterit in written corpus data. The acceptability might, consequently, reflect a mere frequency pattern. Nevertheless, if this is the case, one can argue that this preference might result from an even spread of information, which should facilitate processing. The influence on online processing can, however, not be directly tested with acceptability rating studies but with self-paced reading experiments like the one presented in the next section.

5.2 Experiment 2: Self-Paced reading study

5.2.1 Results

96 participants were recruited through Prolific (Prolific 2024) and needed 18.5 minutes on average to complete the experiment. Eight participants were excluded because their accuracy in answering all comprehension questions, including the fillers, was either below 80% (mean accuracy: 94.4%) or they failed to specify the county in which they had lived the longest.[14] Thus, the data of 88 participants (PMR: n = 55, PLR: n = 33) is analyzed here.

Figure 7 presents the descriptive statistics of the reading times. The critical region, the finite verb, is found in region 2, the spillover region is region 3 and 4, which build the subject NP throughout all items.

[14] The accuracy for the tense items was 97.2% on average. The 'county' specification was given as 'Germany'.

The preterit sentences seem to be read slower than perfect sentences, nearly independent of the sentence onset and the dialect regions. Since the preterit is built with a full verb, its processing effort 'spills over'.[15] Furthermore, differences between the dialect regions become obvious, which were not visible in the acceptability judgement data. The PMR-data (right side of the plot) suggests only an effect of tense but not of the sentence onset in regions 2 to 4, that is, from the finite verb to the subject-NP. In the PLR-condition, however, participants seem to speed up at the subject in the preterit and temporal adverb condition, which would be interpretable in favor of UID.

To test whether these observations are statistically significant, I conducted a linear mixed-effect regression (Bates et al. 2015) in R (R Core Team 2024) for the reading times on the finite verb and the spillover region (determiner and noun of the subject) following Mitchell (2018). The regions were analyzed separately since they were presented word by word. Prior to analyzing the data, reading times under 100 and over 5000 seconds were excluded.

The dependent variables are the log-transformed, centered reading times (logRT). The independent variables are the item position (centered and z-transformed), the sum-coded adverb at the sentence onset (neutral: −0.5, temporal: 0.5), the sum-coded tense (perfect: −0.5, preterit: 0.5), the sum-coded dialect region (PMR −0.5, PLR 0.5) and the length in characters for the

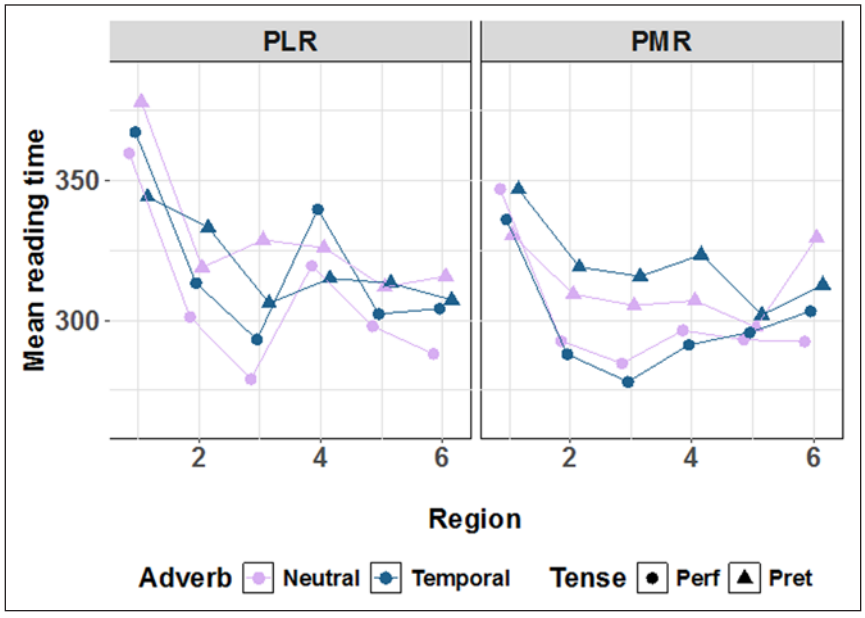

Figure 7: Mean reading times of the sentences. Region 1 is the sentence onset, region 2 the finite verb, region 3 the definite article of the subject-NP, and region 4 the head noun of the subject-NP

[15] I thank a participant from the RAILS conference for this suggestion.

finite verb and the NP, which was also centered and z-transformed. The three-way interaction between tense, adverb and dialect region was also included, as were the random intercepts for the participants for the analysis of the finite-verb-region. For the article- and the noun-region, I included the random intercepts for the items and a random intercept and random slopes for the tense and the participants.

Reading times on the critical region (finite verb)

The results of the linear regression are presented in Table 5.[16] The only significant result is provided by the item position, which is expected, as participants are known to speed up during an experiment when they get used to the task and the procedures. There is also a tendency for preterit to be read slower, but its significance level is above the threshold (t = 1.88, p = 0.6). This result can be attributed to the preterit verb conveying more meaning than the auxiliary in the perfect sentences.

Table 5: Results of the regression analysis, experiment 2. Critical region: finite verb. Random effect: Participant): var = 0.04, std = 0.21. | Onset: var = 0.00008 std = 0.009 | Residual: var = 0.05, std = 0.22

	Estimate	Std. Error	df	t-value	p-value	
Intercept	5.688	0.023	86.58	248.86	<0.001	***
Adverb	0.003	0.011	6.24	0.25	0.80	
Tense	0.042	0.022	1935.94	1.88	0.06	.
Dialect	0.053	0.046	88.18	1.157	0.25	
Length	0.004	0.011	1770.32	0.38	0.71	
Position	−0.094	0.005	2041.71	−19.67	<0.001	***
Adverb:Tense	0.011	0.019	2043.76	0.551	0.58	
Adverb:Dialect	0.007	0.019	2041.03	0.336	0.74	
Tense:Dialect	−0.012	0.019	2040.61	−0.627	0.53	
Adverb:Tense:Dialect	−0.053	0.039	2040.88	−1.358	0.17	

Reading times on the spillover region “article”

Since processing difficulties often become visible after the critical region, this section focuses on the spillover region, the article. The results of the linear regression[17] can be found in Table 6.

[16] lmer(logRT ~ Adverb*Tense*Dialect + Length + Position + (1| Participant)+(1|Onset), region2). The inclusion of the items as random effects instead of the sentence onset resulted in a non-converging model in R. 18 items were excluded here, because the reading times were either below 100 or above 5000 seconds.

[17] lmer(logRT ~ Adverb*Tense*Dialect + Position + (1| Item) + (Tense| Participants), region3).

21 values with reading times either below 100 or above 5000 seconds were excluded.

Again, there is an influence of the item position relating to habituation. Furthermore, the influence of the tense is clearly visible: Participants slow down in preterit sentences ($t = 6.46$, $p < 0.01$). Slower reading times in the preterit condition are also visible in the interaction between tense, adverb and the dialect region (Figure 8). Here, an effect of the sentence onset and the tense is visible for one of the two participant groups.

For the PMR-participants (Figure 8, left side), only the main effect of tense is significant; preterit sentences are read more slowly. The sentence onset does not influence the reading times for this participant group.[18] This changes for the participants from the preterit loss regions (PLR, right side Figure 8). These participants speed up in the preterit condition when a temporal adverb stands at the sentence onset, which is in line with the hypothesis ($t = -1.91$, $p < 0.05$). Spreading the tense information facilitates processing, especially for the preterit. In the perfect condition, the sentence onset does not yet influence the reading times of the PLR-participants.

Table 6: Results of the regression analysis, experiment 2. Region: Spillover (article)

	Estimate	Std. Error	df	t-value	p-value	
Intercept	5.65	0.02	88.64	248.41	<0.001	***
Adverb	−0.007	0.01	94.04	−0.72	0.47	
Tense	0.076	0.01	65.64	6.46	<0.001	***
Dialect	0.03	0.05	87.69	0.59	0.56	
Position	−0.08	0.01	2022.87	18.12	<0.001	***
Adverb:Tense	−0.02	0.02	93.92	−0.91	0.36	
Adverb:Dialect	−0.005	0.02	1929.16	−0.27	0.78	
Tense:Dialect	0.005	0.02	86.25	0.22	0.83	
Adverb:Tense:Dialect	−0.09	0.04	1929.17	−2.34	<0.05	*

Table 7: Random effects, experiment 2, spillover region (article)

Group	Variance	Std. Dev.	Correlation
Item	0.0003	0.02	–
Participants	0.042	0.20	–
Tense|Participants	0.003	0.05	0.7
Residual	0.046	0.21	–

[18] The significance level was tested using the sim_slopes-command from the interactions-package in R (Long, 2019): sim_slopes(mod, pred= Adverb, modx =Tense, mod2=Dialect, johnson_neyman=T).

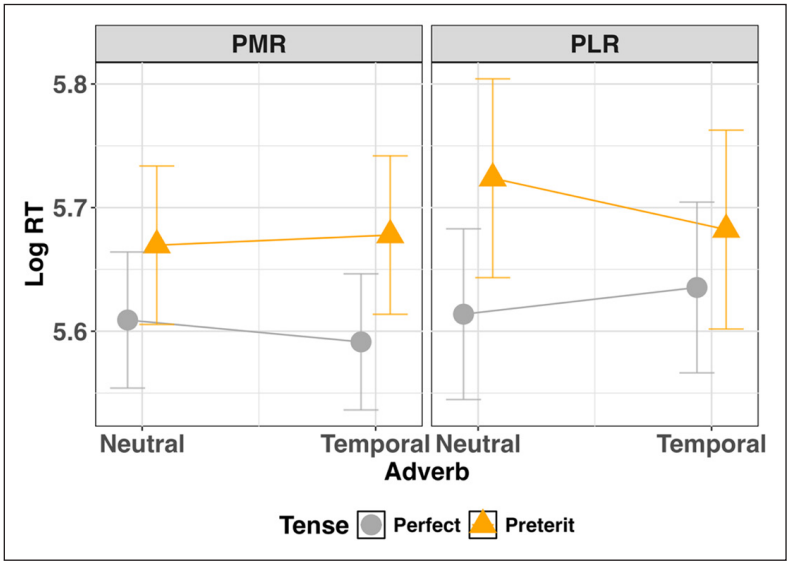

Figure 8: Interaction plot Adverb:Tense:Dialect Region, experiment 2, region 3. The interaction plots are created with ggeffects (Lüdecke 2018) and ggplot (Wickham 2016) in R

Reading times on the spillover region "noun"

The same predictors and the same interaction as before are significant in the second spillover region, i.e., on the noun phrase.[19] Preterit items are generally read slower ($t = 2.62$, $p < 0{,}05$), participants speed up in the course of the experiment ($t = -20.77$, $p < 0{,}001$), longer NPs are read more slowly, and the three-way-interaction is significant ($t = -3.07$, $p < 0.05$) as shown in Table 8 and Figure 9. The interaction of Tense with Dialect Region is now significant as well ($t = -2.50$, $p < 0{,}05$), showing that people from PMR read preterit items faster than perfect, while PLR-people read the two tenses equally fast. In contrast to the results of the article-region, the sentence onset now significantly influences the reading times of the PLR-participants in the perfect items as well: Perfect is read slower in the temporal adverb condition. This is contrary to the hypothesis and will be discussed in the following section.

5.2.2 Discussion

The reading time data for the spillover region suggests that the processing of the preterit is facilitated when it is accompanied by a temporal adverb, compared to a neutral adverb, particularly for participants from regions where the preterit is in decline. This finding supports the hypothesis that the distribution of information across multiple linguistic units can ease cognitive processing, especially in contexts where a given structure, like the preterit, is less

[19] lmer(logRT ~ Adverb*Tense*Dialect+ Position +Length + (1| Item) + (Tense|Participants), region4). 20 values with reading times either below 100 or above 5000 seconds were excluded.

Table 8: Results of the regression analysis, experiment 2. Critical region: spillover (noun)

	Estimate	Std. Error	df	t-value	p-value	
Intercept	5.68	0.03	89.51	202.42	<0.001	***
Adverb	0.001	0.01	95.96	0.11	0.91	
Tense	0.03	0.01	75.22	2.62	0.01	*
Dialect	0.07	0.06	87.96	1.17	0.25	
Length	0.035	0.006	93.30	6.28	<0.001	***
Position	−0.10	0.005	2011.27	−20.77	<0.001	***
Adverb:Tense	−0.02	0.02	95.85	−0.91	0.37	
Adverb:Dialect	−0.007	0.02	1923.38	−0.33	0.74016	
Tense:Dialect	−0.06	0.02	87.38	−2.50	<0.05	*
Adverb:Tense:Dialect	−0.12	0.04	1923.41	−3.07	<0.05	**

Table 9: Random effects, experiment 2, spillover region (noun)

Group	Variance	Std. Dev.	Correlation
Item (Intercept)	0.0007	0.03	–
Participants (Intercept)	0.06	0.25	
Tense|Participants	0.004	0.06	−0.36
Residual	0.048	0.22	–

familiar. That familiarity influences processing is well known (Frank et al. 2016). In the PLR, speakers encounter the preterit less frequently, as it is unlikely that they *hear* it. If at all, they read it and even in written registers, perfect is also used and possibly with more frequency as it can be used in a wider range of contexts.[20] Even in texts that primarily employ the preterit, instances of the perfect are attested, suggesting limited habituation to the preterit form.

Thus, when the preterit appears with a tense-neutral adverb, PLR-participants face a double cognitive strain. Not only are they less accustomed to the preterit itself, but they must also extract all temporal information from a single word rather than distributing it across the verb and the adverb.

[20] Even modern novels are often written in present despite the preterit's propensity for narration (e.g. Meisnitzer 2016). This does not imply that they produce the perfect tense more frequently; rather, it constitutes another instance of the absence of preterit usage.

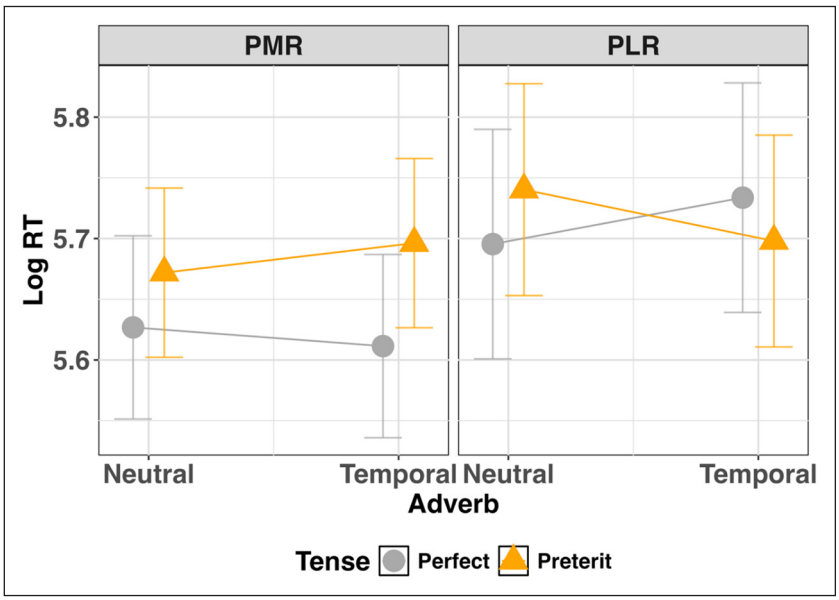

Figure 9: Interaction plot Adverb:Tense:Dialect Region, experiment 2, region 4

Besides providing evidence for hypothesis (H3), it also provides tentative evidence that dialectal variation, such as the preterit loss, can influence the perception and processing of Standard German, highlighting the interplay between regional linguistic habits and cognitive mechanisms in language comprehension. Still, this claim may be too bold here and needs further investigation with a more balanced sample of participants and other phenomena.

It is unlikely that the effect, observed in the article and the noun, is linked to integration difficulties that depend on the finite verb. Even in languages with a rather free word order, such as German, there is a preference for subjects to occur early in the sentence. When they are not placed at the sentence onset, they should therefore appear as soon as possible behind the finite verb and, indeed, they form the next constituent in the sentences. So, at least from a syntactical point of view, there should not be any integration problems. Semantically, I do not expect any problems because the items were designed to be as lexically unobtrusive as possible. Accordingly, I will interpret the effects on those two words as actual spillover effects and, thus, evidence for the hypothesis.

6 General discussion

The current paper aimed to find evidence for the following hypotheses (Section 4.1) using an acceptability judgement task and a self-paced reading experiment.

The acceptability judgment task provides partial support for hypothesis (H1), and full support for its specification in hypothesis (H1.1), as it reveals a significant interaction between tense and sentence-initial adverb type. The sentence onset does not significantly influence the acceptability of perfect tense sentences, but of preterit tense sentences. They are rated higher when they start with a temporal adverb.

I attribute the observed result fully to the Uniform Information Density principle (Levy & Jaeger 2007). UID states that information should be distributed as evenly as possible across utterances to avoid peaks and troughs in the information profile. In the case of preterit sentences, such an unwanted peak is created on the lexical verb in the preterit sentences because participants must process various kinds of information in one word. By presenting words before the finite verb which anticipate some information, the peak can be smoothed out. In this study, the anticipating word is the temporal adverb. Having seen it and expecting grammaticality, recipients know that the verb must be presented in one of the past tenses. As soon as they encounter a full verb instead of an auxiliary, they further know that the tense will be the preterit. In terms of UID, this means that the processing cost of the tense information is spread and, thus, smoothed.

The reason for this effect not being found in perfect sentences is that they already spread parts of the information across two units. The perfect tense is analytic. When recipients encounter the finite auxiliary verb, they basically process all the information they also need to process in preterit sentences – except for the lexical and valency information. Accordingly, the information content of said finite verb is already lower.[21] A counterargument that can be put forward here is that the auxiliaries *haben* 'have' and *sein* 'be' can lead to temporary ambiguities and should, therefore, require a processing effort similar to full verbs in preterit because the auxiliary could be misinterpreted as a homonymous full verb in the present tense (in the neural sentence onset conditions). However, this explanation appears unlikely given the results of a cloze task in the linear, non-cumulative self-paced reading design with the same items. Only 1.56% of all responses indicated a full verb interpretation, and even in the neutral adverb condition, this number only rose to 3.13%. These findings suggest that, while ambiguity may theoretically exist, it does not significantly impact participants' interpretation – at least after having perceived the whole sentence. Whether online processing difficulties arise in the perfect tense conditions (as well) was tested in the self-paced reading study.

Before this second experiment is discussed, I will quickly examine the second hypothesis (H2). Due to the preterit loss in some German regions, it was assumed that the ratings may differ depending on the regions in which the participants have lived longest. However, that was found not to be the case. I attribute this to the fact that the sentences were presented in written form. The preterit

[21] However, not having seen the valency information early in the sentence might result in processing difficulties concerning complements and their order, as described by Voigtmann (accepted).

is linked to the written medium (e.g. Hennig 2000), and even participants from the preterit loss regions are used to the preterit there. Furthermore, in the acceptability judgement task, there is no time pressure for comprehension. It might be interesting to test whether the ratings would differ depending on the stimulus presentation, though this goes beyond the scope of the current study. In addition, the results could also be influenced if the sample of PMR and PLR participants were more balanced or if participants could be found from regions where the preservation or loss of the past tense is particularly pronounced. It would also be interesting to test in further studies to what extent the results presented here change when strong verbs are used instead of weak verbs, which preserve the past tense better. These ideas and concerns also relate to the third hypothesis.

The last hypothesis of this paper (H3) was tested with experiment 2, a self-paced reading study. It was expected that participants would read the finite verb faster in the temporal adverb condition, especially when combined with preterit tense, and that there would be a difference depending on the dialect regions. However, the expected effect was not found on the critical region itself but on the spillover region. On the subject of the sentences, I find evidence for the hypothesis: Participants from the preterit loss regions read both the article and the noun behind the finite verb in the preterit-temporal onset condition significantly faster than in the neutral condition in preterit. The effect is not found for the participants from the preterit maintenance regions. When a tense-neutral adverb precedes the preterit, PLR-participants experience increased cognitive load: they are both less familiar with the preterit and must derive temporal information solely from the adverb, rather than from both verb and adverb. This not only supports the hypothesis but also suggests, albeit tentatively, that dialectal features like preterit loss can shape how Standard German is processed, pointing to an interaction between regional linguistic patterns and cognitive processing. Nevertheless, this interpretation requires further empirical support.

To summarize: I find evidence for hypotheses (H1.1) and (H3), but not for hypothesis (H2), which had the weakest claim in any case.

Despite these results, some open questions remain, especially considering the processing of perfect sentences in experiment 2. The reading time data on the noun-spillover-region indicates that the perfect items were processed more slowly when accompanied by temporal adverbs, suggesting that they require additional processing effort. Tentatively, I want to suggest that the differences between the regions depend on the habituation to the preterit only in written sources as an explanation. Fischer (2018: 135) states in reference to Sieberg (2003) that a temporal adverb increases the likelihood of observing preterit in (written) corpus data. It could be that the PLR-participants do not only expect a past tense but the preterit after reading a temporal adverb. Integrating the perfect in these conditions might cause difficulties. However, further research is needed to relate the findings of this study to this explanation, for example, by changing from a written stimulus presentation to an auditory one, where there should be less habituation to preterit for the people from the PLR.

7 Summary

This paper explores how the UID principle influences the processing of German past tenses – specifically the preterit and the perfect – and how regional dialect variation affects this process. UID predicts that distributing information evenly across an utterance facilitates comprehension by avoiding processing peaks. The preterit, being synthetic, concentrates information in the verb and may cause such peaks, while the analytic perfect (auxiliary + participle) distributes information more evenly. Temporal adverbs (e.g. *gestern*, 'yesterday') introduced early in a sentence help pre-signal tense and support this distribution, unlike tense-neutral adverbs (e.g. *vielleicht*, 'perhaps'). Additionally, the decline of the preterit in southern German dialects (PLR) may reduce speakers' familiarity with this tense.

To find evidence for the hypothesis, two experiments were conducted. The acceptability judgement task found that preterit sentences were rated as more acceptable when preceded by temporal adverbs. No such effect was found for perfect sentences, and no regional differences emerged – likely due to the written format, where the preterit remains common. The self-paced reading experiment shows that, in spillover regions (subject noun phrases), PLR participants read preterit sentences with temporal adverbs significantly faster than those with neutral adverbs. This supports the UID-based prediction that information distribution eases processing, especially when the familiarity with a tense is low. Unexpectedly, PLR participants read perfect sentences with temporal adverbs more slowly, possibly reflecting an expectation for the preterit after such adverbs.

The findings support the UID principle, showing that early temporal adverbs help distribute tense information and reduce processing load – especially for speakers less familiar with the preterit. This highlights how dialectal variation can shape the processing of Standard German. Open questions remain regarding the delayed processing of perfect sentences in PLR regions and the limitations of early region measures in self-paced reading.

8 Data availability, ethics and consent and funding and acknowledgement

The data used for this study is available here:

https://osf.io/9f3c6/overview?view_only=cee85fbb1c4d43f3b02e5c89494ed4be

The reported experiments were conducted as filler experiments in a series of experiments supervized by Bozhidara Hristova (Saarland University) in the context of project B3 in CRC 1102 »Information Density and Linguistic Encoding«. The experiments are covered by the

prolongation (#2021-22-211122) of the ethics lab vote for experimental studies conducted within CRC 1102 (#2017-07-180423), issued by the ethics committee of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS). Before the experiments, participants provided informed consent by checking a checkbox.

The work reported in this paper has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project ID 232722074 – CRC 1102.

I want to thank Bozhidara Hristova and Robin Lemke for their feedback on the experimental design, the audience of the *Rational Approaches in Language Science (RAILS)* conference, where this work has been presented, as well as Prof. Dr. Lea Schäfer and Prof. Dr. Holden Härtl and the two anonymous reviewers for their valuable comments on a previous version of this paper.

References

Abraham, Werner. 2004. The European demise of the simple past and the emergence of the periphrastic perfect: areal diffusion or natural, autonomous evolution under parsing facilitation? In Werner Abraham (ed.), *Focus on Germanic Typology*. 241–272. Berlin: Akademie Verlag. https://doi.org/10.1524/9783050084336.241.

Abraham, Werner. 2005. Präteritumschwund und das Aufkommen des analytischen Perfekts in den europäischen Sprachen: Sprachbundausbreitung oder autonome Entfaltung? In Eckhard Eggers, Schmidt, Jürgen Erich & Stellmacher, Dieter (eds.), *Moderne Dialekte – Neue Dialektologie. Akten des 1. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD) am Forschungsinstitut für deutsche Sprache „Deutscher Sprachatlas" der Philipps-Universität Marburg vom 5.–8. März 2003*. 115–133. Stuttgart: Steiner.

Abraham, Werner & Elisabeth Leiss. (eds.). 2016. *Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik*. Hamburg: Buske.

Ágel, Vilmos & Mathilde Hennig. 2007. Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens. In Vilmos Ágel & Mathilde Hennig (eds.), *Zugänge zur Grammatik der gesprochenen Sprache*. 179–214. Tübingen: Niemeyer (Germanistische Linguistik; 269). https://doi.org/10.1515/9783110936063.179.

Askedal, John Ole. 1986a. Über "Stellungsfelder" und "Satztypen" im Deutschen. *Deutsche Sprache*, 14. 193–223.

Askedal, John Ole. 1986b. Zur vergleichenden Stellungsfelderanalyse von Verbalsätzen und nicht-verbalen Satzgliedern im Deutschen (1). *Deutsch als Fremdsprache*, 23(5). 269–273.

Aylett, Matthew & Alice Turk. 2004 The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, 47(1). 31–56. https://doi.org/10.1177/00238309040470010201.

Bates, Douglas, Martin Mälcher, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Bech, Gunnar. 1955. *Studien über das deutsche Verbum infinitum*. Kopenhagen: Munksgaard.

Christensen, Rune Haubo Bojesen. 2023. Ordinal – Regression Models for Ordinal Data'. https://CRAN.R-project.org/package=ordinal.

Collins, M.X. 2014. Information Density and Dependency Length as Complementary Cognitive Models. *Journal of Psycholinguistic Research*, 43. 651–681. https://doi.org/10.1007/s10936-013-9273-3.

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2). 193–210. https://doi.org/10.1016/j.cognition.2008.07.008.

Díaz-Campos, Manuel & Sonia Balasch. (eds.). 2023. *The handbook of usage-based linguistics*. Hoboken, New Jersey: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119839859.

Divjak, Dagmar. 2019. *Frequency in Language: Memory, Attention and Learning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316084410.

Drinka, Bridget. 2004. Präteritumschwund: evidence for areal diffusion. In Werner Abraham (ed.), *Focus on Germanic Typology*. 211–240. Berlin: Akademie Verlag. https://doi.org/10.1524/9783050084336.211.

Eckardt, Regine & Augustin Speyer. 2014. Information Structure and Language Change. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*. 503–520. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642670.013.23.

Ernst, Thomas. 2009. Speaker-oriented adverbs. *Natural Language & Linguistic Theory*, 27(3). 497–544. https://doi.org/10.1007/s11049-009-9069-1.

Fenk, August & Gertraud Fenk. 1980. Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie*, 27(3). 400–414.

Fischer, Hanna. 2018. *Präteritumschwund im Deutschen: Dokumentation und Erklärung eines Verdrängungsprozesses*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110563818.

Fischer, Hanna. 2021. Präteritumschwund im Deutschen: Neue Erkenntnisse zu einem alten Rätsel. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, 143(3). 331–363. https://doi.org/10.1515/bgsl-2021-0027.

Fleischer, Jürg, Simon Kasper & Alexandra Lenz. 2012. Die Erhebung syntaktischer Phänomene durch die indirekte Methode: Ergebnisse und Erfahrungen aus dem Forschungsprojekt "Syntax Hessischer Dialekte" (SYHD). *Zeitschrift für Dialektologie und Linguistik*, 79(1). 2–42. https://www.jstor.org/stable/41698954.

Frank, Stefan L., Thijs Trompenaars & Shravan Vasishth. 2016. Cross-Linguistic Differences in Processing Double-Embedded Relative Clauses: Working-Memory Constraints or Language Statistics? *Cognitive Science*, 40(3). 554–578. https://doi.org/10.1111/cogs.12247.

Frey, Werner. 2003. Syntactic conditions on adjunct classes. In Ewald Lang, Claudia Maienborn & Cathrine Fabricius-Hansen (eds.), *Modifying Adjuncts*. 163–210. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110894646.163.

Hale, John. 2001. A Probabilistic Earley Parser as a Psycholinguistic Model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*. USA: Association for Computational Linguistics (NAACL '01). 1–8. https://doi.org/10.3115/1073336.1073357.

Hennig, Mathilde. 2000. *Tempus und Temporalität in geschriebenen und gesprochenen Texten*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110929911.

Henriksson, Henrik. 2006. *Aspektualität ohne Aspekt? Progressivität und Imperfektivität im Deutschen und Schwedischen*. Stockholm: Almqvist & Wiksell International.

Hoberg, Ursula. 1973. Vielleicht – wahrscheinlich – sicher. Bemerkungen zu einer Gruppe von pragmatischen Adverbialen. In Hugo Moser (eds.), *Linguistische Studien IV. Festgabe für Paul Grebe zum 65. Geburtstag, Teil 2*. 87–102. Düsseldorf: Schwann.

Höhle, Tilman N. 2018. Der Begriff ‚Mittelfeld': Anmerkungen über die Theorie der topologischen Felder. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle*. 279–294. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.1169673.

Jegerski, Jill. 2014. Self-paced reading. In Jill Jegerski & Bill Van Patten (eds.), *Research methods in second language psycholinguistics*. 20–49. New York: Routledge.

Juzek, Tom S. 2024. Signal Smoothing and Syntactic Choices: A Critical Reflection on the UID Hypothesis. *Open Mind*, 8. 217–234. https://doi.org/10.1162/opmi_a_00125.

Kasper, Simon & Jeffrey Pheiff. 2023. Die Dynamik der (Morpho-)Syntax des Deutschen in der „Vertikale": Perfektexpansion und Abfolge pronominaler Objekte. In Hanna Fischer & Stefan Rabanus (eds.), *Morphologische und syntaktische Variation in den deutschen Regionalsprachen*. 19–74. Baden-Baden: Olms.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe — Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36(1). 15–43. https://doi.org/10.1515/9783110244922.15.

Koch, Peter & Wulf Oesterreicher. 2007. Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik*, 35(3). 346–375. https://doi.org/10.1515/zgl.2007.024.

Latzel, Sigbert. 1977. *Die deutschen Tempora Perfekt und Präteritum: eine Darstellung mit Bezug auf Erfordernisse des Faches 'Deutsch als Fremdsprache'*. München: Hueber.

Leiss, Elisabeth. 1992. *Die Verbalkategorien des Deutschen: Ein Beitrag zur Theorie der sprachlichen Kategorisierung*. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110883541.

Lemke, Robin 2021. *Experimental investigations on the syntax and usage of fragments*. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.5596236.

Levy, Roger. 2008. 'Expectation-based syntactic comprehension', *Cognition*, 106(3). 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hofmann (eds.), *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*. 849–856. The MIT Press. https://doi.org/10.7551/mitpress/7503.003.0111.

Löbner, Sebastian. 1989. German *schon – erst – noch*: an integrated analysis. *Linguistics and Philosophy*, 12(2). 167–212.

Löbner, Sebastian. 2002. Is the German Perfekt a perfect Perfect? In Graham Katz, Sabine Reinhard & Philip Reuter (eds.), *Sinn & Bedeutung VI, Proceedings of the Sixth Annual Meeting of the Gesellschaft für Semantik*. 255–273. Osnabrück: University of Osnabrück. https://doi.org/10.18148/sub/2002.v6i0.831.

Long, Jacob A. 2019. *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*. https://doi.org/10.32614/CRAN.package.interactions.

Lüdecke, Daniel. 2018. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, 3(26), 772. https://doi.org/10.21105/joss.00772.

Maienborn, Claudia & Martin Schäfer. 2011. Adverbs and adverbials. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science*. 1390–1420. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110255072.1390.

Manning, Christopher D. & Hinrich Schütze. 2005. *Foundations of Statistical Natural Language Processing*. 8. Cambridge, Mass.: MIT Press.

Meisnitzer, Benjamin. 2016. *Das Präsens als Erzähltempus im Roman: eine gedruckte Antwort auf den Film*. Tübingen: Narr.

Meister, Clara, Tiago Pimentel, Patrick Haller, Lena Jäger, Ryan Cotterell & Roger Levy. 2021. Revisiting the Uniform Information Density Hypothesis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic: Association for Computational Linguistics. 963–980. https://doi.org/10.18653/v1/2021.emnlp-main.74.

Mitchell, D.C. 2018. An Evaluation of Subject-Paced Reading Tasks and Other Methods for Investigating Immediate Processes in Reading 1. In David E. Kieras & Marcel A. Just (eds.) *New Methods in Reading Comprehension Research*. 69–90. London: Routledge. https://doi.org/10.4324/9780429505379-4.

Musan, Renate. 2002. *German Perfect: Its semantic composition and its interactions with temporal adverbials*. Dordrecht: Springer.

Naveed, Humza, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes & Ajmal Mian. 2023 A Comprehensive Overview of Large Language Models. arXiv. https://doi.org/10.48550/ARXIV.2307.06435.

Niebaum, Hermann & Jürgen Macha. 2014. *Einführung in die Dialektologie des Deutschen*. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110338713.

Nübling, Damaris. 2016. Die nicht-flektierbaren Wortarten. In Angelika Wöllstein et al. (eds.), *Der Duden in zwölf Bänden: das Standardwerk zur deutschen Sprache*. 9., vollständig überarbeitete und aktualisierte Auflage. 567–633. Berlin: Dudenverlag.

Ogihara, Toshiyuki. 2006 Tense, adverbials and quantification. In Raffaella Zanuttini, Héctor Campos, Elena Herburger & Paul H. Portner (eds.), *Crosslinguistic research in syntax and semantics: Negation, Tense, and Clausal Architecture*. 231–247. Washington, D.C.: Georgetown University Press.

Ogihara, Toshiyuki. 2011. Tense. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics. An International Handbook of Natural Language Meaning. Vol. 2*. 1463–1484. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110255072.1463.

Ortmann, Katrin, Sophia Voigtmann, Stefanie Dipper & Augustin Speyer. 2024. An information-theoretic account of constituent order in the German middle field. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*. 55–86. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.13383787.

Prolific. 2024. 'Prolific'. London, UK: Prolific. https://www.prolific.com, 23. February 2026.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/, 23. February 2026.

Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. London: Collier-McMillan.

Reis, Marga. 1980. On Justifying Topological Frames : "Positional Field" and the Order of Nonverbal Constituents in German⁰. *DRLAV. Documentation et Recherche en Linguistique Allemande Vincennes*, 22/23. 59–85. https://doi.org/10.3406/drlav.1980.957.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1. 75–116. https://doi.org/10.1007/BF02342617.

Rothstein, Björn. M. 2006. The perfect time span : on the present perfect in German, Swedish and English. Stuttgart: Universität Stuttgart. https://doi.org/10.18419/OPUS-2614.

Sapp, Christopher D. 2009. Syncope as the Cause of *Präteritumschwund*: New Data from an Early New High German Corpus. *Journal of Germanic Linguistics*, 21(4). 419–450. https://doi.org/10.1017/S1470542709990134.

Schäfer, Lisa. 2021. Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft*, 40(2). 161–197. https://doi.org/10.1515/zfs-2021-2024.

Schäfer, Lisa. 2025. The licensing and usage of topic drop in German. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.15584708.

Schäfer, Martin. 2013. *Positions and Interpretations: German Adverbial Adjectives at the Syntax-Semantics Interface*. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110278286.

Schallert, Oliver. 2014. *Zur Syntax der Ersatzinfinitivkonstruktion: Typologie und Variation*. Tübingen: Stauffenburg.

Scharlau, Ingrid, Ulrich Ansorge & Odmar Neumann. 2003. Spezielle Verfahren IV: Reaktionszeitmessung: Grundlagen und Anwendung. In Gert Rickheit, Theo Herrmann & Werner Deutsch (eds.), *Psycholinguistik. Psycholinguistics: Ein internationales Handbuch*. 190–202. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110114249.2.190.

Sichelschmidt, Lorenz & Elena Carbone. 2003. Experimentelle Methoden. In Gert Rickheit, Theo Herrmann and Werner Deutsch (eds.), *Psycholinguistik. Psycholinguistics: Ein internationales Handbuch*. 115–124. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110114249.2.115.

Steinhauer, Karsten & Michael T. Ullman. 2002. Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language, 83(1)*. 62–65.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *The Bell Sytsem Technical Journal*, 27(3). 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Sieberg, Bernd. 2003. Regelhafte und normale Anwendung von Perfekt und Präteritum: Mit Anregungen für den DaF-Bereich. *Jahrbuch der ungarischen Germanistik*. 291–315.

Specht, Larissa. 2023. The time course of adverbial oder processing – Psycholinguistic studies on adjunct processing. Tübingen: Universität Tübingen. https://doi.org/10.15496/PUBLIKATION-79922.

Speyer, Augustin. 2008. German vorfeld-filling as constraint interaction. In Anton Benz & Peter Kühnlein (eds.), *Constraints in Discourse*. 267–290. Amsterdam, Philadlphia: John Benjamins Publishing Company. https://doi.org/10.1075/pbns.172.13spe.

Speyer, Augustin. 2009. Das Vorfeldranking und das Vorfeld-es. *Linguistische Berichte*, (219). 323–354. https://doi.org/10.46771/2366077500219_4.

Speyer, Augustin. 2010. Filling the Vorfeld in spoken and written discourse. In Sanna-Kaisa Tanskanen, Marja-Liisa Helasvuo, Marjut Johansson & Mia Raitaniemi (eds.), *Discourses in Interaction*. 263–290. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.203.19spe.

van der Auwera, Johan. (ed.) 1998. *Adverbial Constructions in the Languages of Europe*. De Gruyter Mouton. https://doi.org/10.1515/9783110802610.

Vater, Heinz. 1994. *Einführung in die Zeit-Linguistik*. Hürth-Efferen: Gabel.

Voigtmann, Sophia. accepted. The Object Order in the German Middle Field through the Lens of Information Theory: A Diachronic Study, in: *Journal of Germanic Linguistics*.

von Polenz, Peter. 2013. *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart; Band II, 17. und 18. Jahrhundert*. 2. Auflage. Edited by Claudine Moulin and Dominic Harion. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110314670.

Wall, Joanna & Sophia Voigtmann. 2025. New Perspectives on the Order of Pronominal Objects in New High German. SaRDiS 12 – Saarbrücker Runder Tisch für Dialektsyntax, Saarbrücken [Presentation].

Wallenberg, Joel C., Rachael Bailes, Christine Cuskley & Anton K. Ingason. 2021. Smooth Signals and Syntactic Change. *Languages*, 6(2). 60. https://doi.org/10.3390/languages6020060.

Weiß, Helmut. 1998. *Syntax des Bairischen: Studien zur Grammatik einer natürlichen Sprache*. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110912487.

Welke, Klaus. 2005. *Tempus im Deutschen: Rekonstruktion eines semantischen Systems*. Berlin, Boston: de Gruyter. https://doi.org/10.1515/9783110925371.

Wenker, Georg. 1888. *Sprachatlas des Deutschen Reichs. Handgezeichnetes Original*. Marburg. https://regionalsprache.de/wa.aspx, 23 February 2026.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. https://ggplot2.tidyverse.org, 23 February 2026.

Wiesinger, Peter. 1983. Die Einteilung der deutschen Dialekte. In Werner Besch, Ulrich Knoop, Wolfgang Putschke & Herbert E. Wiegand (eds.), *Dialektologie, Part 2*. 807–900. Berlin, New York: de Gruyter. https://doi.org/10.1515/9783110203332.807.

Wöllstein, Angelika. 2014. *Topologisches Satzmodell*. 2. Edition. Heidelberg: Winter.

Wöllstein, Angelika & Bibliographisches Institut (eds.). 2022 *Duden: die Grammatik*. 10., völlig neu verfasste Auflage. Berlin: Dudenverlag (Der Duden in zwölf Bänden, Band 4).

Zehr, Jérémy & Florian Schwarz. 2022. 'PennController for Internet Based Experiments (IBEX)'. https://doi.org/10.17605/OSF.IO/MD832.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 3

https://doi.org/10.18148/zs/2026-2003

Dawei Jin, Shanghai Jiao Tong University, China, daweijin@sjtu.edu.cn

Jun Chen, University of Stuttgart, Germany, junchen113@gmail.com

Wide Scope and Interpretation of *why*: Experimental Evidence for Causal Interrogatives

Abstract: This paper reports three experiment studies investigating the unique syntactic and interpretive properties of the why-adjunct in Mandarin Chinese in comparison to other wh-types. The empirical findings are: (i) Mandarin why-questions demonstrate a different intervention effect from other wh-questions; (ii) multiple wh-sentences with the causal interrogative phrase paired with additional interrogative phrases yield interpretations that are notably less natural than those with other wh-phrases; and (iii) why-questions lack non-interrogative uses, such as free-choice or indefinite meanings. These results converge in showing that the causal adjunct exhibits an exceptional wide scope, consistent with one prominent line of theoretical research.

Keywords: *why*-question, intervention effect, multiple *wh*-question, non-interrogative meaning

Cite as: Jin, Dawei & Jun, Chen. 2026. Wide Scope and Interpretation of *why*: Experimental Evidence for Causal Interrogatives. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi.org/10.18148/zs/2026-2003
Received: 06 November 2024; **accepted:** 28 January 2026; **published:** 25 June 2026

1 Introduction

A longstanding assumption posits that the causal why-adjunct behaves differently from other wh-phrases (Lawler 1971; Cattell 1978; Bromberger 1992; Hegarty 1992; Iatridou & Kroch 1992; Oshima 2007; Tomioka 2009; Jin 2019; Soare 2017; 2021).[1] The core claim is that the why-constituent always has a wide scope reading, tied to its proposition-modifying semantics (Rizzi 1990; 2001; Bromberger 1992; Murphy 2017). Specifically, 'why' is base-generated at the licensing position of the question-typing Q operator, which is above the TP-internal position where other wh-phrases are assumed to merge, based on the idea that the licensing of C-level categories takes place after the licensing of T-level operators.

The exceptional wide scope of 'why' has been diagnosed by its resistance to semantic embedding. Notably, 'why' is never interpreted within the scope of another operator (Lawler 1971; Cattell 1978; Bromberger 1992). The English *why*-question in (1b) must receive the interpretation where *why* stays outside the quantified proposition:

(1) Why did at most a dozen students register for this class?
   a. 'At most a dozen students registered for this class. Why is that?' (*why* > *at most*)
   b. *'What reason is such that at most a dozen students register for this class because of that particular reason?' (**at most* > *why*)

(1b) is compatible with the situation in which no more than a dozen (among the context-relevant students) registered for class. It is nonetheless not compatible with the reading where the situation contains several groups of students that each registered for the class for a different reason, and the speaker seeks a reason that would account for a group of no more than a dozen students.

The restriction to only a wide scope reading is not limited to a particular syntactic position (e.g. the subject position). As (2a)–(2b) further show, *why* cannot be interpreted within the scope of adverbs, either (modified from Tomioka 2009; Jin 2019).

(2) a. Why is Adam **often** on vacation?
   'Adam is often on vacation. Why?' (*why* > *often*)
   *'What reason is such that Adam goes on vacation mostly because of that particular reason?' (**often* > *why*)
   b. Why is Adam **likely** to accept this offer?
   'Adam is likely to accept this offer. Why?' (*why* > *likely*)
   *'What reason is such that it is likely that Adam accepts this offer because of that particular reason?' (**likely* > *why*)

[1] Throughout this paper we use 'why' as a generic representative of the causal adjunct independent of language. We use the italic form *why* to refer to the causal adjunct specifically in English.

A related high attachment proposal is put forward in Rizzi (2001) for *perché* 'why' in Italian, motivated by its interaction with focus. Rizzi argues that a regular wh-phrase is associated with the position of focus projection below CP, compatible with the fact that the co-occurrence of wh-constituents and contrastive focus is degraded.[2] In contrast, as (3) shows, *perché* is able to appear together with contrastive focus (indicated via capitalization) and crucially, co-occurrence *only* takes place when *perché* c-commands the focused constituent (not the other way round), which is indicative of a wide scope interpretation.

(3) a. Perché QUESTO avremmo dovuto dir-gli, non qualcos'altro?
why this have.FUT.1PL must.PST.PTCP say-3SG.DAT not something.else
'Why THIS we should have said to him, not something else?'

b. *QUESTO perché avremmo dovuto dir-gli, non qualcos'altro?
this why have.FUT.1PL must.PST.PTCP say-3SG.DAT not something.else
#'THIS why we should have said to him, not something else?'

Hence the empirical picture points to a relation of causal modification taking place at the clausal level, distinct from other modification at the verbal level.

On the basis of a series of acceptability rating tasks, this paper presents experimental evidence from Mandarin Chinese for the claim of the uniqueness of 'why'. Mandarin *weishenme*, along with similar why-adjuncts in other East Asian languages like Japanese and Korean, has traditionally been at the core of proposals within the external merge, high attachment approach to the why-adjunct (Ko 2005; 2006; Tsai 2008; Stepanov & Tsai 2008; Tomioka 2009; Takita & Yang 2014; Endo 2015; Jin 2016; Miyagawa 2017; Murphy 2017; Cheng 2021). This body of work provides novel, language-specific insights that, until now, have not undergone systematic testing. Taken together, they could build a strong case for a wide scope interpretation. Three testable diagnostics developed in these accounts will be of particular interest to us here. First, 'why' is argued to trigger an intervention violation that is more robust than that induced by the rest of wh-phrases. Second, 'why' cannot be interpreted coherently in multiple wh-questions, in contrast to other wh-phrases. Third, 'why' is uniquely excluded from expressing a range of non-question readings such as universal or existential readings.

We believe controlled experiments play an irreplaceable role in confirming the above generalizations. This is because for the arguments about 'why' being unique by way of its idiosyncratic syntactic, semantic and pragmatic features to work, it is crucial to establish that the judgments for 'why' in the relevant contexts (e.g. an intervening-violating configuration) are qualitatively different from those for the rest of wh-types. The issue is made subtle, however, by the fact that wh-phrases are not homogeneous. Adjuncts *in general* introduce degradation

[2] Rizzi shows that the focus projection in Italian is restricted to the focus expression that overtly expresses a contrast, realized by a stress pattern.

in acceptability judgments, following from properties that distinguish them from arguments, e.g. their inability to anchor an individualizable domain, be referential or be discourse-linked (d-linked). It is thus crucial to be certain that whatever pattern obtained is indeed attributable to the exceptional nature of 'why' rather than a broader argument-adjunct difference. To this end, quantitative methodology is desirable with its ability to detect multiple levels of contrast (Gibson & Fedorenko 2010; Gibson et al. 2013; Schütze 2016; Linzen & Oseki 2018), compared to the more limited scope of introspection-based, judgment-of-minimal-pairs analyses.

The novel contribution in this paper thus lies in obtaining a clear understanding of whether and to what extent the judgment of why-sentences differs from that of other wh-types by conducting an experimental evaluation. Specifically, we conducted three rating-based formal acceptability tasks, using stimuli representing intervention-violating configurations, multiple wh-configurations and non-interrogative contexts.

The rest of this paper is structured as follows. Section 2 presents the relevant background. Section 3 reports the experiments. Section 4 discusses the findings. Section 5 concludes the paper.

2 Theoretical background

In this section, we discuss three diagnostics from previous literature that make the case for the exceptional wide scope of 'why' in Mandarin. Specifically, these diagnostics compare 'why' with other wh-phrases in contexts of intervention effects, multiple wh-questions, and non-interrogative wh-sentences. We then subject these diagnostics to empirical testing by conducting three experiments, which are detailed in Section 3.

2.1 Intervention effects

The distinction between 'why' and other wh-phrases factors into the pattern of intervention effects. We understand wh-intervention in its standard sense as a constraint wherein a wh-restriction cannot have a scopal operator other than its own question operator as its closest c-commanding operator (Hoji 1985), i.e. (4).

(4) *[Q [Op [… *wh* …]] …]

(5) instantiates one such garden-variety intervening configuration, because a scopal element stays closer to the wh-phrase than its Q operator.

(5) ??Zhiyou tamen jia maidao-le shenme?
only they family buy-PRF what
'What did only their family successfully purchase?' *[Q_j [F-Op_i [… F_i …] […*what*$_j$ …]]]

Importantly, while all wh-types are observed to incur a degraded judgment to a certain degree under the intervening configuration, they do not pattern the same. Rather, the literature has often characterized the causal adjunct in Mandarin as inducing the most robust/reliable/severe intervention effect (Tsai 1994; Cheng & Rooryck 2000; Kim 2002; Ko 2005; Soh 2005; Yang 2011; Li 2011; Jin 2016). According to these reported judgments, the following structure with 'why' being c-commanded by a focus phrase is more degraded compared with a corresponding intervening configuration featuring other wh-phrases.

(6) *Zhiyou tamen jia weishenme maidao-le youpiao?
only they family why buy-PRF stamp
Intended: 'For what reason$_i$ did only their family successfully purchase the stamps t_i?'

A predominant approach where this is accounted for relies on the notion that wh-adjuncts are less acceptable in intervention contexts than wh-arguments, and 'why' is particularly so by way of being the paradigmatic adjunct (Cheng & Rooryck 2000; Law 2001; Kim 2002; Soh 2005; Yang 2011). This approach encounters a problem in Japanese and Korean, where the causal adjunct presents another exceptional case to the generalization for wh-intervention yet the pattern is quite opposite to the case in Mandarin. Whereas Mandarin 'why' induces particularly robust intervention, 'why' in Japanese and in Korean are uniquely *exempt* from intervention (Ko 2005; Tomioka 2009). Take (7) for illustration. Here the why-adjunct *naze* is situated in an intervention-violating configuration (it is c-commanded by a focus expression). The structure is nevertheless judged as acceptable, even though other wh-phrases are not acceptable in a comparable configuration.

(7) Daremo naze ko-nak-atta-no? (Japanese)
anyone why come-NEG-PST-Q?
'Why did no one come?'

Both the pattern of 'why' in Mandarin and the one in Japanese and Korean receive a unified explanation, if (instead of classing 'why' as a prototypical wh-adjunct) we adopt the approach where 'why' takes scope differently (Ko 2005; Tomioka 2009; Li 2011; Jin 2016; 2019). In this view, 'why' is attached high, its base position being where it is interpreted, i.e. the scope position of the Q operator. Now a focus phrase needs to be interpreted in the scope of a sentence type/speech act operator such as Q. It thus must also stay below the scope of 'why', as in (8). In other words, even with 'why' at its base position, we arrive at a configuration that induces no intervention, given that said base position is high enough.

(8) [Q_j *why*$_j$ [F-Op_i [… F_i …]]]

This approach allows us to account for the circumvention of 'why' in Japanese and Korean as in (7). We can do so, if we further consider the scrambling characteristic of both languages:

Scopal elements in Japanese and Korean can undergo a free word order operation that affects surface order (i.e. scramble). Although focus may precede Q at the surface level of representation, this happens via scrambling, which is standardly taken to apply on the PF side of the grammar and thus does not modify the LF structure. At LF, the focus operator still scopes below Q. This way, the surface form of (7) has an LF as in (8), and it is the latter that feeds into the semantic representation, giving rise to an unproblematic interpretation.

The unacceptability of 'why' in Mandarin also follows, assuming one difference with Japanese and Korean: Scopal elements in Mandarin are strictly isomorphic (Huang 1982; Ernst 1994; Huang et al. 2009). They cannot scramble. A surface structure with focus c-commanding 'why' such as (6) would require an LF as in (9), assuming scope isomorphism. However, this is uninterpretable.

(9) *[Zhiyou$_{\text{F-Op}}$ [tamen jia]$_{\text{F}}$ [$_{\text{Q}}$ weishenme maidao-le youpiao?]]
only they family why buy-PRF stamp
Intended: 'For what reason$_i$ did only their family successfully purchase the stamps t_i?'

⇝ *[F-Op$_i$ [… F$_i$ …] [Q$_j$ *why*$_j$…]]

When taking account of one parametric variation, the wide scope proposal therefore enables us to account for both the ungrammaticality in Mandarin and the circumvention in Japanese/Korean.

Crucially, the wide scope explanation comes down to a property specific to 'why'. If this explanation holds, the ungrammaticality induced by Mandarin 'why' will be an altogether different phenomenon from the garden-variety wh-intervention represented above in (4), despite their apparent similarity (cf. Jin 2022). Recall that canonical cases of wh-intervention feature a focus operator *scoping between* Q and its wh-restriction. Differing from that, the ungrammaticality represented by the structure in (9) is induced, because there can be no higher scopal operator c-commanding the already high-attached 'why'. There is simply no such position available for the interpretation of focus in Mandarin. Said ungrammaticality is triggered without real 'intervening', as no separation between Q and its wh-restriction is involved. Operating from the view that 'why' does not trigger *bona fide* intervention, the wide scope approach to 'why' thus predicts that acceptability judgment of 'why' differs significantly from that of canonical wh-adjuncts in an intervening context in Mandarin. Alternatively, if 'why' belongs with other wh-adjuncts in terms of position and interpretation (the why-as-adjunct view), it is predicted that acceptability judgment for why-sentences patterns with that of canonical wh-adjuncts.

2.2 Multiple *wh*-questions

Another aspect where the uniqueness of 'why' manifests itself lies in its inability to give rise to a coherent interpretation in a multiple wh-question. A Mandarin constituent question is formed with the wh-phrase staying *in situ*, i.e. occupying the same position as its non-interrogative counterpart. The wh-in-situ property applies, both when the question contains a single wh (e.g. 10), and when multiple wh-s are involved (e.g. 11).

(10) Ni zai na'er xia dan de?
you LOC where place order PRT
'Where did you place your order?'

(11) a. Ni shenme shihou zai na'er xia dan de?
you what time LOC where place order PRT
'When and where did you place your order?'
b. Ni shenme shihou zai na'er zenme xia dan de?
you what time LOC where how place order PRT
'When, where, and how did you place your order?'

Notably, *weishenme* 'why' is argued to stand out, excluded from a multiple-wh reading that is elsewhere available to other wh-phrases (Stepanov & Tsai 2008; Takita & Yang 2014).[3] This is exemplified in (12).[4]

(12) a. *Ni weishenme shenme shihou xia dan de?
you why what time place order PRT
Intended: 'Why and when did you place your order?'
b. *Ni weishenme shenme shihou zai na'er xia dan de?
you why what time LOC where place order PRT
Intended: 'Why, when and where did you place your order?'

The wide scope approach to *weishenme* renders an explanation for its lack of multiple-wh readings (Takita & Yang 2014; Endo 2015; Jin 2019; 2020). Consider first how a single-wh interpretation proceeds. As (13) illustrates, a regular wh-restriction is assumed to be base-generated at the T domain, where it introduces a variable (wh_i) at one of the argument slots of the proposition it is contained in ([$_{TP}$… wh_i…]). At a higher level of compositional interpretation,

[3] One anonymous reviewer points out that the degraded judgment appears to be ameliorated with the position of *shenme shihou* swapped against that of *weishenme*. In our following multiple-wh experiment, we adopted this supposedly rescuing configuration to see the proposed generalization about the exclusion of *weishenme* from a multiple-wh reading can still be maintained. We return to this issue briefly in Section 4.

[4] Here the literature traditionally distinguishes between a single-pair reading (e.g. (11a) expects a unique time-location pair as answer) and a pair-list reading (e.g. (11a) asks that for each individual time, what is the corresponding location) (Engdahl 1986; Dayal 2017). Both interpretations are unavailable to (12).

this propositional radical combines with a C head (i.e. C^0). For a question reading to come about, the C head is to be endowed with a [+wh] feature. Following Dayal (2017), what this C head does semantically is to take in said proposition and outputs a set of propositions, crucial to the question interpretation. Finally, at the level above C head, a question operator Q_i at the specifier of the C domain evaluates this set of alternatives.[5]

(13) $[_{\text{Spec, CP}}\ Q_i\ [_{C'}\ i\ [_{C'}\ C^0\ [+wh]\ [_{TP}\ wh_i\ ]]]]$

In the presence of multiple wh-restrictors, each will need to be evaluated by a separate Q operator. A multiple-wh interpretation may arise, if we assume that all question operators merge simultaneously at the specifier of the C-head (syntactically speaking they undergo absorption). Thus, the multiple-wh structure in (11a) (repeated below as 14a) receives an LF as in (14b).

(14) a. Ni <u>shenme shihou</u>$_i$ <u>zai na'er</u>$_j$ xia dan de?
you what.time LOC where place order PRT
'When and where did you place your order?'

b. $[_{\text{Spec, CP}}\ Q_i\ Q_j\ [_{C'}\ j\ [_{C'}\ i\ [_{C'}\ C^0\ [+wh]\ [_{TP}\ wh_i\ wh_j\ ]]]]]$

A proposition containing two variables is fed to the C head (wh_i and wh_j), which yields a set of open propositions. Composition within the C-domain proceeds in parallel to a single wh-question, barring that there is a recursive application of the process where a Q operator composes with a lambda-abstracted open proposition introduced by an index (Q_i with i and Q_j with j).[6]

Consider now the multiple-wh configuration with 'why' paired up with another wh. With 'why' evaluated at an exceptional wide scope above that of a regular wh, the above mechanism cannot apply. The difference is that 'why' does not leave behind a variable at one of the argument slots of the proposition radical. Rather, it directly resides where the Q operator is and is thus above the C head (Takita & Yang 2014). This yields the LF as in (15).

(15) LF for (12a): $[_{\text{Spec, CP}}\ [Q_j\ j\ why_j]\ Q_i\ [_{C'}\ i\ [_{C'}\ C^0\ [+wh]\ [_{TP}\ \ldots wh_i \ldots]]]]$

(15) cannot be interpreted like (14b), and moreover, cannot give rise to any interpretations due to an interpretive conflict: Because the other wh-restriction is below C, the C head needs to be [+wh]-endowed. It takes in a proposition, and outputs a set of propositions which is then fed to 'why' at the level above C. However, here a conflict arises with regards an independent

[5] In keeping with Dayal's system, we assume that this process is done via lambda abstraction: An index introduces lambda abstraction of the wh-restriction that bears that index (e.g. the index i in (13)). Afterwards, a Q operator composes with the lambda-abstracted open proposition introduced by an index, which introduces quantification over the variable. The formal details are not crucial.

[6] In the pair-list reading the two indices may enter into a functional relation via Skolemization as in Dayal (2017), to be ignored here.

truthfulness presupposition. This presupposition, projected by *and only by* why-questions, states that the content expressed by the modified result clause is true (e.g. Hintikka & Halonen 1995: 647).[7] In our case, the modified result clause of 'why' is a set of propositions (a question, following the Hamblin/Karttunen sense), which by definition is not evaluable with a truth value. Consequently, an interpretation failure ensues. Hence, 'why' is only compatible with a single-wh question (where the C head is not [+wh]-endowed), but must be excluded from multiple-wh questions.

For this analysis to be viable, we need to be certain that 'why' uniquely resists multiple-wh contexts. This is necessary in order to rule out other possibilities. A most obvious one is to consider 'why' as an instance of adjunct wh-phrases being incompatible with a multiple-wh environment. There is independent evidence that this is the case in at least some wh-in-situ structures. In English, for example, the judgment of a multiple-wh configuration may be off with both *how* and *why* staying *in situ* (e.g. Huang 1982; Huang et al. 2009).[8]

(16) a. *Who fainted how?
b. *Who arrived why?

The pattern has been addressed by assuming that wh-adjuncts do not have a set of individuals (or do not easily anchor a domain of individuals from context). We will not go into detail to make the discussion overcomplex here (cf. e.g. Reinhart 1997). Suffice it to say that while the exact pattern in multiple wh-in-situ is not settled, the issue becomes relevant in deciding between accounts. Specifically, we expect to observe in Mandarin a uniquely severe degradation for 'why'

[7] Following the prevalent view since Belnap (1969), a why-question expresses causal modification by relating a proposition (i.e. a cause) to another proposition (i.e. the modified result) (see also Szabolcsi & Zwarts 1993; Hintikka & Halonen 1995; Schwarz & Simolenko 2018). For the causation relation to obtain in the first place, the result itself has to be a given, hence the content expressed by the result clause must be true (Hintikka & Halonen 1995; Tomioka 2009; Larson & Sawada 2012; Schwarz & Simolenko 2018; Fortmann 2019; Zifonun et al. 1997: 2292ff.). Thus, the why-question in (1b), repeated below, asks for the reason why the proposition that at most a dozen students registered for class is true.

(i) Why did at most a dozen students register for this class?
Presuppose: At most a dozen students registered.

Based on the same intuition, 'why' has been treated as selecting for a fact (Mellor 1995; 2004), which is ontologically separate from a proposition. The distinction is not crucial for the current account. Note that this special relation holds independent of the temporal specifications of the clauses involved. This truthfulness presupposition may be conceived as a reflex of the relation that is established between the causal explanation and its result: Causal dependence is characterized by a necessity relation, such that assuming there is a cause/explanation, then its result *must* follow as a consequence. Further assuming that wh-questions in general project an existential presupposition, then it follows that the question *why-p* presupposes that there exists a cause for *p*, from the domain of 'why'. From this we derive the truthfulness presupposition.

[8] The generalization with English how-in-situ is not straightforward. While (16a) is degraded, *who arrived how?* is seen as more toward the acceptable end. Thus rather than rigid ungrammaticality, the pattern of judgments shows significant variation, depending on lexical choices.

in a multiple wh-question, under an account that explains the ungrammaticality of 'why' via its high attachment. On the other hand, with 'why' treated as a garden-variety adjunct similar to 'how', it is predicted that 'why' should pattern with 'how' under multiple wh-configurations.

2.3 Non-interrogative *wh*-sentences

Another behavior in which 'why' has been argued to behave differently from the rest of wh-phrases involves its lack of non-interrogative uses: As in many languages (Kuroda 1965), wh-items in Mandarin can express the readings of universals, existential indefinites, free choice items (FCIs) or negative polarity items (NPIs), yet these functions noticeably exclude 'why'. We focus on two of these uses here, the use of wh as an indefinite and as a free choice item.

Wh-indefinites are widely attested across languages (Haspelmath 1997). Mandarin belongs with a class of languages in exhibiting the so-called bare wh-indefinite use, where indefinites are identical in form with the corresponding interrogatives without introducing additional morphology (Yun 2019). This is in contrast to another class of languages, including Japanese and Korean, where wh-indefinites are derived from interrogatives by attaching an affix. In following with a generalization over bare indefinite languages, wh-indefinite expressions in Mandarin are limited to a narrow scope reading, rather than taking free scope as regular indefinites (Cheng 1991; Li 1992; Lin 1998; 2004; Bruening 2007; Yun 2019). They also differ from regular indefinites by having a non-specific and non-referential interpretation. The following instances of wh-indefinite use are only valid when the speaker does not have a witness to the wh-expression in mind. This obligatorily non-specific reading is often reinforced by the co-occurrence of evidentials like *haoxiang* 'apparently' or illocution-weakening particles like *yixia*, which helps with disambiguation by excluding an interrogative reading (Lü 1942; Chao 1968; Xue 2014).

(17) a. Ta shou-li haoxiang ti-le ge shenme dongxi.
she hand-inside seem carry-PRF CLF what thing
'She appears to be carrying something in her hands.'

b. Jiu kan ta zenme caozuo-le yixia, wo de diannao jiu you neng yong-le.
just see her how operate-PRF one.bit I POSS computer then again can use-PRF
'We just watched her operating for a bit on that computer somehow, and my computer was back in business again.'
(taken from Chinese Web 2017 (zhTenTen17) Simplified, Jakubíček et al. 2013)

It has been shown that *weishenme* 'why' differs from other wh-phrases and cannot be interpreted non-interrogatively in an environment that otherwise licenses an indefinite use, as illustrated in the unacceptable case in (18) (Shi 2021; Niu 2023):

(18) *Benlai qingjia-le, ta haoxiang weishenme you keyi lai-le.
originally ask.for.leave-PRF she seem why after.all can come-PRF
Intended: 'She had originally asked to be off, yet it seems like she can come in after all for some reason.'

Previous research has also observed a free choice use that is derived from wh-interrogative words in Mandarin (Lü 1942; Cheng 1991; Giannakidou & Cheng 2006; Sugimura 2007). The free choice reading comes about, when the wh-element is followed by an universalizing morpheme, such as the distributive operator *dou* and *ye*, with *ye* restricted to negative contexts. Thus, the following corpus-attested examples cannot be interpreted as a question, but rather receive an *any*-like interpretation:

(19) a. Shei dou xihuan ta.
who DOU like her
'Anyone would like her.'
b. Zhiyao xin li you sha, na'er dou shi maerdaifu.
as.long.as heart inside have sand where DOU COP Maldives
'As long as there is a place in the heart for sand, anywhere could be the Maldives.'
(taken from Chinese Web 2017 (zhTenTen17) Simplified, Jakubíček et al. 2013)

The example in (20) is taken to show that *weishenme* 'why' also fails to license a free choice reading, contra other wh-phrases (Shi 2021; Niu 2023).

(20) *Yuangong weishenme bei jiegu dou hui dedao san-bei peichangjin.
employees why PASS sack DOU will receive three-times severance.pay
Intended: 'Employees will be given three times the normal severance pay, regardless of the reason for their dismissal.'

To derive non-interrogative interpretations, it is proposed that the Hamblin alternatives expressed by the wh-phrase are evaluated by different embedding operators, instead of being evaluated by the Q operator (Kratzer & Shimoyama 2002; Giannakidou & Cheng 2006; Beck 2019). This means that in wh-indefinites, an operator introduces existential quantification over the Hamblin alternatives. Universal quantification applies to the set of alternatives in the case of a free choice reading. Once again, this process could be subject to an asymmetry by adjuncthood. The gist is that an adjunct that lacks an internal structure is inherently less amenable to anchoring a closed/enumerable domain of individuals, rendering it unclear what set of alternatives the embedding operator is applied to. It could then follow that 'why' lacks non-interrogative interpretations due to its adjunct status (Sugimura 2007; Xue 2014; Shi 2021; Niu 2023).

Alternatively, deviating from subsuming 'why' under adjuncts, Jin (2016; 2020) and Cecchetto & Donati (2017) argue that a wide scope analysis explains the idiosyncratic distribution of 'why'. Note first that the embedding (existential/universal) operators are assumed to bind into the restrictor set of a wh-phrase below the C domain. With the wh-variable now existentially/universally bound, the C head no longer composes with an open proposition and derives a set of propositions out of it for the question interpretation at higher levels (cf. section 2.2 for the role of C head in wh-interrogatives). Instead, it simply selects a propositional radical with no free variables, like it does with any declarative-type sentence radicals. Crucially, however, the way such non-interrogative reading comes about relies on the wh-variable staying within the domain of these embedding operators. Meanwhile, if we assume that the wh-restrictor of 'why' merges exceptionally high and attaches to the scope of the Q operator, it then follows that the Hamblin alternatives of 'why' never interact semantically with the embedding operators. This way, the only option is for the why-variable to receive binding by the Q operator, making an interrogative reading the only possible interpretation.

In a nutshell, we once again arrive at two contrasting predictions, depending on the view adopted. As with previous contexts, a wide scope view predicts that participants will assign a distinctly low judgment rating to why-sentences in non-interrogative contexts. This prediction is in contrast to the one from the alternative, why-as-adjunct view, which groups 'why' with other wh-adjuncts such as 'how', and therefore predicts comparable rating patterns between these adjunct types.

In sum, suffice it to say that there has been a body of observations pointing to the exceptional nature of Mandarin 'why'. These pertain to a distinctive pattern in intervention violations, as well as the unavailability of multiple-wh and non-interrogative interpretations. These patterns, if established, offer evidence for treating the why-adjunct as being attached high where its scope is resolved. At the same time, experimentally testing these claims using rigid formal methods is needed to make sure that the empirical basis of the characterization of 'why' is as solid as possible, which is particularly so given the non-homogeneous nature among wh-types. We now turn to our new study, which aims to fill in a noticeable empirical lacuna by experimentally investigating the behavioral idiosyncrasies of 'why' across three environments.

3 Experiment

The theoretical implications of the current paper crucially follow from the empirical claim that 'why' is unique. To test this claim, we conducted three formal judgment surveys on the following environments: (1) focus intervention-violating configurations, (2) multiple wh-constructions, (3) non-interrogative wh-constructions (including wh-indefinites and wh-free choice items). The

goal is to see if acceptability ratings on a Likert scale differ significantly by wh-type. In the following, we report our experiments in turn. The stimuli lists used in the three experiments of our survey research, as well as a description of the complete statistical analysis and the R code scripts, are available in the OSF repository.

A note on our choice of wh-types and use of the term 'adjunct' is in order here. It is long noted that East Asian wh-in-situ phrases receive different licensing mechanisms. This observation dates back to Huang (1982), who initially proposed a division based on the familiar distinction between syntactic arguments and adjuncts. Subsequent work on wh-in-situ has nevertheless shown that this is not the most revealing contrast. A more theoretically-relevant distinction is nominal status, i.e. whether a wh-element contains a transparent, divisible nominal component. On this dimension, expressions like *shenme shihou* 'what time' or *shenme difang* 'what place' are adverbial but nominal, as they include a decomposable morphological argument within an otherwise adverbial projection. In contrast, adverbials like *zenme* 'how' and *weishenme* 'why' (on its causal reading) are non-nominal. They do not break down into isolable parts. It is argued that only nominal phrases introduce an individual variable, the right type of variable that can be bound at a distance, whereas non-nominals, which cannot supply an individual variable, necessarily trigger feature movement (Tsai 1994; 1999; 2008; Stepanov & Tsai 2008). Tsai and collaborators thus argue that this distinction modulates acceptability judgments in configurations that involve wh-in-situ and covert movement.

Against this backdrop, the current design considers three predictions: (1) 'why' patterns with any other wh-adjuncts, with no distinction whatsoever between them, or (2) 'why' patterns with non-nominal adjuncts like 'how', or (3) 'why' behaves in a *sui generis* way, distinct from both nominal and non-nominal adjuncts. Hypothesis (1) is included only as a minimal baseline, and the substantive theoretical contrast lies with the other two predictions. To decide between these predictions, in the experiments we chose a nominal adjunct, *shenme shihou* 'what time', to contrast with *zenme* and *weishenme*, the latter two being non-nominal. Such choice also has the benefit of ensuring positional consistency. If we adopt a DP like *shenme* 'what' to serve as the nominal baseline, it would not match the positions of a wh-adverbial. In using 'what time' (instead of 'what'), we preserve the predictive adequacy of the nominality theory and simultaneously remove a potential confound arising from positional mismatches.

3.1 Experiment 1

Competing approaches seek to account for the intervention behavior of 'why'. As mentioned above, we mainly consider two substantive approaches. The first approach treats 'why' as a typical, non-nominal wh-adjunct, attributing its intervention behavior to this status. We define

nominality in terms of whether a wh-expression has an individual(ized) domain, realized as a divisible, transparent internal structure. The temporal adverb *shenme shihou* '(literally) what time' has an internal structure on a par with wh-DPs, whereas the manner adverb *zenme* does not. Given 'what time', 'how' and 'why', this view predicts that 'why' does not differ significantly from 'how' in intervention-violating environments. If we observe any difference between 'why' and 'what time', that difference should be explained by a categorial distinction along the line of nominality, i.e. we expect to observe a comparable difference between 'how' and 'what time'. In contrast, the second approach treats 'why' as unique among wh-phrases due to its wide scope, a property not shared by other wh-elements. This approach posits that the intervention effects associated with 'why' differ from canonical wh-intervention. Given 'what time', 'how' and 'why', it predicts that any difference between 'why' and 'what time' cannot be explained in terms of the difference between 'how' and 'what time'.

Participants A total of 50 individuals from Shanghai Jiao Tong University, a public university in China, participated in the experiment. Our language background survey revealed that all participants were self-reported native speakers of Mandarin Chinese. None reported significant prior exposure to a language beyond Mandarin before age 14. Subjects were each compensated with 20 yuan RMB for their participation. We adopted a filler accuracy threshold of 85%.[9] Eight participants were excluded from analysis for falling below this threshold. Forty-two participants remained in the analysis, comprising 16 males and 26 females (age 25.33±4.26).

Materials The experimental stimuli followed a 3×2 factorial design. Three types of wh-expressions were tested: *weishenme* 'why', *zenme* 'how' and *shenme shihou* 'what time'. Among them, 'how' was chosen to contrast with 'why' as a canonical non-nominal wh-phrase, and 'what time' stood in as a baseline for nominal wh-phrases. These were crossed with a focus phrase consisting of constituent 'only' (*zhiyou* in Mandarin) and its focus associate DP. Crosslinguistically and in Mandarin, focus phrases are reported to give rise to the most robust intervention effect (Kim 2002; Soh 2005; Beck 2006; Yang 2011), which we discuss further in section 4. One sample set of target stimuli is given as follows.

(21) a. *Weishenme* 'why'-question

Zhiyou xuesheng weishenme mai tiyu caipiao?
only student why buy sports lottery.ticket
'Why do only students buy sports lottery tickets?'

[9] Our design included both grammatical and ungrammatical fillers. We treated an ungrammatical filler response as erroneous if it received a score of 4 or above, based on a 7-point Likert scale. Similarly, we treated a grammatical filler response as erroneous if it received a score of 4 or below. We then calculated the overall filler accuracy rate for a given participant.

b. *Zenme* 'how'-question

Zhiyou xuesheng zenme mai tiyu caipiao?
only student how buy sports lottery.ticket
'How do only students buy sports lottery tickets?'

c. *Shenme shihou* 'what time'-question

Zhiyou xuesheng shenme shihou mai tiyu caipiao?
only student what time buy sports lottery.ticket
'When do only students buy sports lottery tickets?'

In our design we opted for a nominal adverbial over a DP (e.g. *shenme* 'what'). This is due to the fact that the latter would occupy the object position as in (22), a consequence of Mandarin being wh-in-situ with SVO order. The problem is that choosing a DP introduces a potential confound for lack of a match in position with 'how' and 'why' in the other conditions.

(22) *Shenme* 'what'-question

Zhiyou xuesheng mai shenme?
only student buy what
'What do only students buy?'

Testing the adverbial 'what time' alongside 'how' thus allows us to probe whether or not 'why' can be subsumed under existing formulations of adjuncthood, while simultaneously eliminating a potential confound introduced by positional mismatches.

In choosing to address the concern with positional confound, our design differed from the one in the recent experiment study of Mandarin intervention in Jin & Yan (2024), which tested wh-DPs. Note in addition that Jin & Yan (2024) probed a much broader empirical scope, contrasting focus phrases against quantifiers of varying monotone classes to determine the robustness of each individual intervener type. Our study adopted a narrowly focused approach, settling upon one intervener type with the aim of a larger sample size and increased statistical power for the exclusive analysis of the behavior of 'why'.

To gauge whether an intervention effect is there, we additionally tested the corresponding controls. Contrary to an intervention-violating target sentence, which has a focus phrase occupy a structurally higher position than the wh-phrase, the c-commanding phrase in the control sentence is a definite plural, which is widely seen as not introducing a structural or interpretational violation in that position (Kotek 2014; 2017; Li & Law 2016). (23) illustrates a sample set corresponding to (21). The c-commanding DP consists of an NP selected by the definiteness marker *zheixie* 'these', the latter a combination of a demonstrative form *zhei* and a pluralized classifier *xie*.[10]

[10] We thank the anonymous reviewer for suggesting the use of the current control design.

(23) a. *Weishenme* 'why'-question control

Zheixie xuesheng weishenme mai tiyu caipiao?
DEM.PL student why buy sports lottery.ticket
'Why do these students buy sports lottery tickets?'

b. *Zenme* 'how'-question control

Zheixie xuesheng zenme mai tiyu caipiao?
DEM.PL student how buy sports lottery.ticket
'How do these students buy sports lottery tickets?'

c. *Shenme shihou* 'what time'-question control

Zheixie xuesheng shenme shihou mai tiyu caipiao?
DEM.PL student what time buy sports lottery.ticket
'When do these students buy sports lottery tickets?'

Including controls thus enabled us to obtain the participants' baseline acceptability for the sentences with an intervening configuration: The structure in (21) gives rise to an intervention effect if a difference approaching statistical significance is observed between the judgment acceptability of said structure and its control structure as in (23).

Target sentences were categorized into three lists and assigned to three groups of participants in a Latin square design. Each participant rated 27 sentences in the main trial. These included nine test stimulus sentences with different lexical items. Stimulus items were pseudo-randomly interspersed with twice as many distractor fillers. These fillers were evenly divided into those that were chosen to be apparently acceptable and those that were apparently unacceptable. Control items followed the same Latin square arrangement, yielding another three lists. With both target and control stimulus items included, 378 data points were sent to statistical analysis.

Procedures The experiment was executed in an online environment using the Tencent Wenjuan survey platform (https://wj.qq.com/).[11] After providing informed consent and completing a preliminary language background survey and familiarization with the instructions, participants commenced the tasks. Each task featured a set of four practice stimuli prior to the introduction of the main trials.

Participants assessed the sentences on a 7-point Likert scale (ascending order, 1 being the least acceptable). Sentences were presented individually on a screen. Participants read a sentence and

[11] As a popular survey platform in China, Tencent Wenjuan served solely as the online hosting interface for our questionnaire. Participants were recruited externally via advertisement postings as well as through social networks, which were then provided with a link directing them to the survey. The platform offered a stable, user-friendly interface familiar to China-based respondents. The platform's features suffice to fulfill our purpose of eliciting Likert-based rating judgments. While it lacks certain advanced functionalities, such as recording response times, these are not essential to our current design.

then indicated their choice by clicking on the corresponding point along a scale underneath that sentence, before proceeding to the next sentence on the following page. The task averaged 290 seconds. Note that all sentences were in Chinese characters. The samples presented here have been transcribed in pinyin Romanization for exposition.

Results Given that an interval-scaled data type was being treated here, and because different participants may use these scales inconsistently, each participant's raw responses were transformed using a z-score procedure (Sprouse et al. 2011; 2012; 2013; Schütze & Sprouse 2014). At the same time, we retained the untransformed (raw) Likert ratings in the analysis to ensure that the direction and magnitude of the key contrasts were not artifacts of the z-score transformation. The main text of this paper reports results based on the untransformed scores. Complete analyses of both data types (z-scored and untransformed) are provided in the online repository. Importantly, the z-score transformation preserved all relationships present in the original data and introduced no distortion.

Figure 1 shows the rating results after z-score transformation.

Table 1 lists model-based point estimates for z-scored ratings by type (target vs. control) and condition, with means (SD) and 95% confidence intervals.

Visual inspection of the rating pattern suggests that control conditions, which instantiate an intervention-free c-commanding scopal expression, were rated better than target conditions across three wh-expressions. For the target conditions, 'what time' and 'how' showed similar distributions, whereas 'why' was rated noticeably lower. We additionally ran an ordinal mixed model (contrast coded against the 'how' condition, Christensen 2019) on the rating data. The model (specified as a cumulative link mixed model, `clmm (Rating ~ wh.category * type + (1|participant) + (1|item))`) predicted ratings from a fixed effect of `wh.category` interacting with `type` (`target` or `control`) and two random effects: By-`participant` and by-`item`. Post hoc pairwise comparison was conducted using the `emmeans` package with Tukey's Honest Significant Difference test (with α-adjustments). The model revealed a significant main effect of `type`, such that target sentences received significantly lower ratings than their corresponding controls. In addition, there was a significant interaction between `wh.category` and `type`, indicating that the effect of wh-category differed between target and control groups. To unpack this interaction, we separately conducted pairwise comparisons within the target and the control conditions. Within the target condition, sentences with 'why' were rated significantly lower than those with 'how' ($\hat{\beta}$: −1.45±0.35, $p < 0.001$) and 'what time' ($\hat{\beta}$: −1.87±0.35, $p < 0.0001$). No significant difference was observed between sentences with 'how' and with 'what time'. Within the control condition, ratings did not differ significantly across wh-types, indicating that wh-category had no detectable effect on the acceptability of control items.

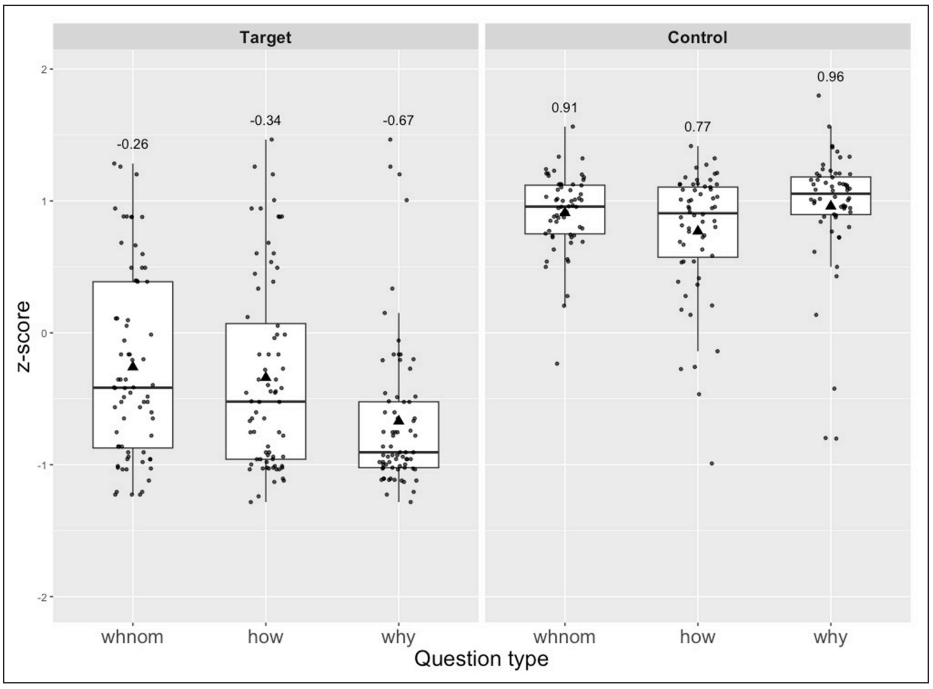

Figure 1: Acceptability judgment ratings across wh-type in the intervention configuration with standardized z-score transformation, represented in a combined boxplot and scatterplot. The left panel represents the target, while the right panel represents the control. In each panel, the three boxes correspond to the three question types: *what time* (coded as `whnom`), *how*, and *why*. Transformed z-scores range from −1.28 to 1.80. Numbers on the plot (as well as the triangles within the respective boxes) represent the calculated means. Horizontal solid lines on the boxes represent median values. The closer the absolute value to zero, the less it deviates from conditioned standard deviation. Positive and negative signs of the numbers represent the direction of a deviation.

Table 1: Z-scored ratings in Experiment 1.

Condition type	Wh-type	z-scored means (SD)	95% CI
Target	whnom	−0.26 (0.71)	[−0.50, −0.02]
Target	how	−0.34 (0.73)	[−0.58, −0.10]
Target	why	−0.67 (0.58)	[−0.91, −0.43]
Control	whnom	0.91 (0.29)	[0.70, 1.13]
Control	how	0.77 (0.48)	[0.55, 0.99]
Control	why	0.96 (0.46)	[0.74, 1,18]

Overall, this experiment has uncovered a significant difference between why-questions and other types of wh-questions with respect to focus-induced intervention. Why-questions elicited more pronounced intervention effects, while a convergence in judgments was observed between 'how' and 'what time'.

An anonymous reviewer wondered about the factor of the possible readings of 'why' and its effect on the acceptability judgment. 'Why' is known to have ambiguous interpretations. Thus, *weishenme* may ask for the cause of an action as well as the purpose of one (the latter based on a *wei* 'for' + *shenme* 'what' parse). Importantly, on the purposive parse *weishenme* is structurally on a par with other nominal wh-phrases, which is expected to lead to an amelioration of the intervention violation. To the extent where purpose 'why' is available to the speakers, its rescuing reading would therefore bias acceptability *upwards*. This means our observed degradation reported above makes for a conservative estimate of how unacceptable causal 'why' is: If participants obtained a mixed interpretation during the task, those obtaining a strictly causal interpretation should assign an even lower score. Our claim about the distinctiveness of 'why' can therefore be maintained when interpretive ambiguity is considered. In fact, the current pattern could even underestimate the effect size of the distinctive behaviors of causal 'why'.

To better understand how interpretive ambiguity factors in boosting acceptability judgments, we conducted a follow-up analysis.[12] We were able to do so, by taking advantage of the fact that the stimuli divided into two (equal-numbered) groups of lexical sets. One group lexically disfavors a purposive reading, whereas the other group allows it. The first, 'exclusive' group featured nonhuman-denoting lexicalizations (e.g. *why do only woodpeckers eat insects?*), where the actions are naturally interpreted causally rather than purposively, because such agents lack full intentional agency and goal-oriented mental representation. In contrast, the second, 'ambiguous' group used human-denoting lexicalizations (e.g. *why do only students buy lottery tickets?*), where the actions can plausibly be understood both in causal terms and in terms of teleologically guided behaviors. Thus, if our experiment participants indeed obtained a purposive reading for *weishenme* in the 'ambiguous' group, and if purposive readings were inflating ratings, we would expect an interaction between *wh*-category and group, that is, compared with the other wh-items, the difference between 'ambiguous' and 'exclusive' *weishenme* is bigger, with 'ambiguous' *weishenme* receiving relatively higher ratings than 'exclusive' *weishenme*.

[12] Another approach brought to bear on this issue is to situate target wh-questions within an appropriate preamble that interpretively biases to one of the two available readings. We did not opt to pursue this approach for the workload it involves; supplying contextualization in a rigid experimental paradigm would lead to a full implementation across all condition types as well as across experiments, going beyond the scope of the current study. See a related analysis in the recent work in Jin & Yan (2025), which restricts the possible readings by supplying a context that settles for one reading per wh-type.

We therefore fitted a mixed-effects model including *wh*-category, group (exclusive vs. ambiguous) and their interaction, and compared it to a simpler model without the interaction. Our model comparison showed no improvement when the interaction was added. In other words, group (reading) did not modulate the pattern across *wh*-categories, that is, the difference in rating between the two groups for 'why' is not bigger than, say, for 'what time'. We take this preliminary result to indicate that, at least for the current experiment, the availability of purposive reading does not significantly impact acceptability judgment. We believe the most likely explanation is that while *weishenme* allows itself to be interpreted ambiguously, not all participants actually obtained a purposive parse of *weishenme*, hence the lack of a significant finding. Alternatively, it is possible that our participants indeed obtained the two available parses, yet the acceptability judgments corresponding to them were not distinct enough to reach statistical significance. Note independently that Kim et al. (2023)'s experiment yielded a finding towards this direction. It would be interesting to explore both directions further and afterwards to find the underlying reasons for them.

In sum, we believe the precise impact of ambiguous interpretation (how big an effect it has on acceptability judgment) remains an open empirical question that needs further investigation. At the same time, to the extent that a purposive reading of *weishenme* is available, it would, if anything, be expected to raise acceptability ratings rather than lower them. Taken together, the evidence is compatible with causal 'why' being less acceptable than the other wh-types in focus intervention contexts.

3.2 Experiment 2

We proceed to our second acceptability judgment task, designed to test whether multiple interrogative constructions involving the why-adjunct exhibit an idiosyncratic behavior. Again, the substantive approaches we consider are as follows. One approach treats 'why' as a typical, non-nominal wh-adjunct. Given 'what time', 'how' and 'why', it predicts that multiple wh-questions with 'why' should pattern similarly to those with 'how'. Both should elicit a comparable decrease in acceptability judgment relative to 'what time'. A second approach treats 'why' as uniquely exhibiting wide scope. It follows from the wide scope property that 'why' cannot yield a coherent interpretation when combined with another wh-element to form a multiple wh-question. Consequently, this approach predicts that the decrease in acceptability observed for 'why' (relative to 'what time') cannot be explained in terms of a similar difference observed between 'how' and 'what time'.

Participants A total of 128 individuals from Shanghai Jiao Tong University, a public university in China, participated in this experiment. The data from eight participants was excluded from analysis for falling below our filler accuracy threshold of 85%. The data of 120

participants was included in the subsequent analysis (42 males and 78 females, 25.14±3.77). Our language background survey revealed that all participants were self-reported native speakers of Mandarin. None reported significant prior exposure to a language beyond Mandarin before age 14. Subjects were each compensated with 10 yuan RMB for their participation.

Materials We created three conditions of target sentences, with `wh.category` manipulating three types of phrases: *weishenme* 'why', *zenme* 'how', and *shenme shihou* 'what time'. In addition to the wh-category type in question, each multiple interrogative question contained another controlled/unvaried wh-phrase *shui* 'who'. The resulting three target conditions are each illustrated with one sample stimulus in (24).

(24) a. Multiple wh-question with *weishenme* 'why'
Shui weishenme bangjia-le fushang Huo xiansheng?
who why kidnap-ASP rich.businessman Huo(Surname) Mister
'Who kidnapped Mr. Huo the business tycoon why?'

b. Multiple wh-question with *zenme* 'how'
Shui zenme bangjia-le fushang Huo xiansheng?
who how kidnap-ASP rich.businessman Huo(Surname) Mister
'Who kidnapped Mr. Huo the business tycoon how?'

c. Multiple wh-question with *shenme shihou* 'what time'
Shui shenme shihou bangjia-le fushang Huo xiansheng?
who what time kidnap-ASP rich.businessman Huo(Surname) Mister
'Who kidnapped Mr. Huo the business tycoon when?'

We also created another three conditions of control sentences across the three *wh*-categories to serve as a baseline of acceptability.[13] These are single wh-sentences that correspond with the target multiple-wh questions, with the unvaried wh-phrase replaced by a non-interrogative DP. A sample set of the control stimuli is given in (25).

(25) a. Single wh-question with *weishenme* 'why'
Nimen weishenme bangjia-le fushang Huo xiansheng?
2PL why kidnap-ASP rich.businessman Huo(Surname) Mister
'Why did you guys kidnap Mr. Huo the business tycoon?'

b. Single wh-question with *zenme* 'how'
Nimen zenme bangjia-le fushang Huo xiansheng?
2PL how kidnap-ASP rich.businessman Huo(Surname) Mister
'How did you guys kidnap Mr. Huo the business tycoon?'

[13] We owe it to an anonymous reviewer for suggesting that we contextualize our target data against appropriate control items in Experiment 2, as well as in Experiment 3 which is to follow.

c. Single wh-question with *shenme shihou* 'what time'
Nimen shenme shihou bangjia-le fushang Huo xiansheng?
2PL what time kidnap-ASP rich.businessman Huo(Surname) Mister
'When did you guys kidnap Mr. Huo the business tycoon?'

Altogether there were nine target sentences, which were then distributed across three groups using a Latin Square design, so that participants saw a unique lexical item in each target condition. Each participant rated three target sentences and six filler sentences. Control items followed the same Latin square arrangement, yielding another three lists. In total, 360 data points entered the statistical analysis.

Procedures The procedures of Experiment 2 were the same as Experiment 1. The stimulus items were pseudo-randomly interspersed with twice as many fillers serving as distractors. The entire experiment, including the background survey and the task, lasted around 4 minutes.

Results **Figure 2** shows the rating results by condition after z-score transformation. Key point estimates from Experiment 2 are provided in **Table 2**.

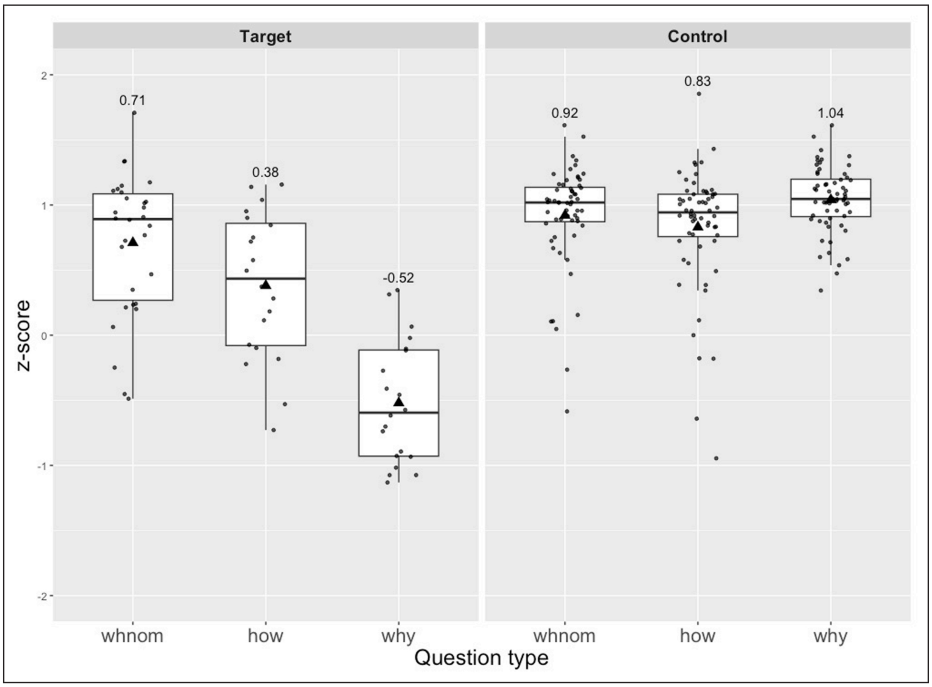

Figure 2: A box plot, augmented with a scatterplot, showing the z-score transformed ratings across wh-type pertaining to the multiple-wh paradigm (left panel: target; right panel: control).

Table 2: Z-scored ratings in Experiment 2.

Condition type	Wh-type	z-scored means (SD)	95% CI
Target	whnom	0.71 (0.54)	[0.39, 1.04]
Target	how	0.38 (0.56)	[−0.01, 0.78]
Target	why	−0.52 (0.47)	[−0.91, −0.11]
Control	whnom	0.92 (0.40)	[0.78, 1.07]
Control	how	0.83 (0.48)	[0.68, 0.97]
Control	why	1.04 (0.25)	[0.90, 1,19]

An inspection of the rating distributions showed a clear difference between target and control items. In the multiple-wh (target) context, 'why' sentences received the lowest rating, followed by 'how' and 'what time' in that order. In contrast, the control sentences exhibited a uniform pattern: all three wh-categories were rated consistently high with no visible pairwise differences, and participant ratings were clustered closely together.

Our ordinal mixed model (`clmm(Rating~ wh.category * type + (1|participant)+ (1|item))`) was fitted to predict Likert scores from a fixed effect of `wh.category` (contrast coded with the 'how' condition) interacting with `type` (`target`, `control`), as well as by-`participant` and by-`item` random effects. Our models revealed a significant main effect of `type`, such that target sentences received significantly lower ratings than their corresponding controls. In addition, there was a significant interaction between `wh.category` and `type`, indicating that the effect of wh-category differed between target and control items. To unpack this interaction, we separately conducted pairwise comparisons within the target and the control conditions. Specifically, sensitivity to wh-category distinctions was present only in the target group. Within the target conditions, post hoc comparisons showed that sentences containing 'why' were rated significantly different from 'what time' conditions ($\hat{\beta}$: −3.51±1.02, $p < 0.001$), as well as 'how' conditions ($\hat{\beta}$: −3.06±0.95, $p \approx 0.003$). The 'what time' condition was rated higher than the 'how' condition, with the difference being marginal. In contrast, the post hoc analysis of the control data did not reveal significant differences among any pairs of these three conditions. Overall, the current experiment revealed that multiple wh-questions involving the why-adjunct are generally unacceptable, in contrast to multiple wh-questions involving 'how' and 'what time'.

3.3 Experiment 3

We now proceed to our third acceptability judgment task, with the aim of testing whether non-interrogative constructions involving the why-adjunct behave differently from those involving

'how' and 'what time'. We conducted two sub-experiments, with the first one testing wh-indefinites and the second one free choice wh-items. In both environments, we test the wide scope approach to 'why' against an alternative, why-as-adjunct approach. Given 'what time', 'how' and 'why', we look into whether 'why' elicits a decrease in acceptability judgment relative to 'what time', and if so, whether said difference is comparable to a similar decrease elicited by 'how'. The approach treating 'why' as a canonical non-nominal predicts that the behaviors of 'why' and 'how' pattern together. In contrast, the approach treating 'why' as uniquely exhibiting wide scope predicts a distinctive, severely more degraded acceptability pattern for 'why'.

Participants A total of 127 participants from Shanghai Jiao Tong University, a public university in China, were recruited for the experiment. The data from 7 participants was excluded from analysis for falling below our filler accuracy threshold of 85%. The data of 120 participants entered statistical analysis (45 males, 75 females, 25.84±3.59). Our language background survey revealed that all participants were self-reported native speakers of Mandarin. None reported significant prior exposure to a language beyond Mandarin before age 14. Subjects were each compensated with 10 yuan RMB for their participation.

Materials We conduced two subexperiments, one for the wh-indefinite paradigm and one for the wh-free choice item paradigm. For each subexperiment, we created three conditions of target sentences, which varied by wh-category. Three types of wh-phrases were manipulated: *weishenme* 'why', *zenme* 'how', and *shenme shihou* 'what time'. 'How' was introduced as a standard case of adjunct against which the behaviors of the why-adjunct were compared, and 'what time' served as a baseline for nominal wh-phrases. A sample set of target stimuli in the indefinite environment is given in example (26).

(26) a. Wh-indefinite construction with *shenme shihou* 'what time'

Dajia dou bu gen xiaoming shuohua. Guji ta shenme shihou
everyone DOU NEG with Xiaoming talk by.estimate he what time
dezui-guo ren.
offend-EXP person
'No one's talking to Xiao Ming. Guess he ruffled some feathers at some point.'

b. Wh-indefinite construction with *zenme* 'how'

Dajia dou bu gen xiaoming shuohua. Guji ta zenme
everyone DOU NEG with Xiaoming talk by.estimate he how
dezui-guo ren.
offend-EXP person
'No one's talking to Xiao Ming. Guess he somehow ruffled some feathers.'

c. Wh-indefinite construction with *weishenme* 'why'

Dajia dou bu gen xiaoming shuohua. Guji ta weishenme dezui-guo ren.
everyone DOU NEG with Xiaoming talk by.estimate he why offend-EXP person
'No one's talking to Xiao Ming. Guess he for some reason ruffled some feathers.'

We additionally created matching control sentences across wh-types in order to establish a baseline of acceptability. The gist was to include a similar question form, this time with a truly interrogative interpretation. A sample set of the control stimuli is given in (27). To enforce an interrogative reading and to exclude a non-interrogative one, a sentence adverb *daodi* 'in the world, on earth' replaced *guji* 'by estimate', which modifies a question speech act (Ernst 1994; 2001). A question-final interrogative particle *ne* is further appended to go with the *daodi*-adverb and ensure that the sentence is overtly marked as the question type (Cheng 1991).

(27) a. Wh-interrogative construction with *shenme shihou* 'what time'

Dajia dou bu gen xiaoming shuohua. Ta daodi shenme shihou dezui-guo ren ne?
everyone DOU NEG with Xiaoming talk he on.earth what time offend-EXP person Q
'No one's talking to Xiao Ming. When in the world did he ruffle some feathers?'

b. Wh-interrogative construction with *zenme* 'how'

Dajia dou bu gen xiaoming shuohua. Ta daodi zenme dezui-guo ren ne?
everyone DOU NEG with Xiaoming talk he on.earth how offend-EXP person Q
'No one's talking to Xiao Ming. How in the world did he ruffle some feathers?'

c. Wh-interrogative construction with *weishenme* 'why'

Dajia dou bu gen xiaoming shuohua. Ta daodi weishenme dezui-guo ren ne?
everyone DOU NEG with Xiaoming talk he on.earth why offend-EXP person Q
'No one's talking to Xiao Ming. Why in the world did he ruffle some feathers?'

A sample set of stimuli in the free choice environment is given in example (28).

(28) a. Wh-free choice construction with *shenme shihou* 'what time'

Wo juede ba, ta name lei, shenme shihou chu cuo-le dou keyi
I feel PRT she that tired what time commit mistake-ASP DOU can
yuanliang.
forgive
'I think given that she's so tired, she can be forgiven no matter when she has made a mistake.'

b. Wh-free choice construction with *zenme* 'how'

Wo juede ba, ta name lei, zenme chu cuo-le dou keyi
I feel PRT she that tired how commit mistake-ASP DOU can
yuanliang.
forgive
'I think given that she's so tired, she can be forgiven no matter how she has made a mistake.'

c. Wh-free choice construction with *weishenme* 'why'

Wo juede ba, ta name lei, weishenme chu cuo-le dou keyi
I feel PRT she that tired why commit mistake-ASP DOU can
yuanliang.
forgive
'I think given that she's so tired, she can be forgiven no matter what reason caused her to make a mistake.'

Wh-question forms with a truly interrogative reading were included as the control, illustrated as in (29).

(29) a. Wh-interrogative construction with *shenme shihou* 'what time'

Ni juede, ta shenme shihou chu cuo-le ne?
you feel she what time commit mistake-ASP Q
'When do you think has she made a mistake?'

b. Wh-interrogative construction with *zenme* 'how'

Ni juede, ta zenme chu cuo-le ne?
you feel she how commit mistake-ASP Q
'How do you think has she made a mistake?'

c. Wh-interrogative construction with *weishenme* 'why'

Ni juede, ta weishenme chu cuo-le ne?
you think she why commit mistake-ASP Q
'Why do you think has she made a mistake?'

While both wh-indefinites and wh-free choice constructions are known to carry special prosody distinct from the interrogative use (Hu 2002; Sugimura 2007), we believe that given that we are dealing with a reading-based task, extra caution is sensible to make sure that participants are truly obtaining a non-interrogative reading for the target sentences and that there is no ambiguity in the available interpretation. To do so, in this design we created a comprehension question for each of the test items: Upon finishing a test sentence, the participant read on the subsequent page a dialogue of the following form in example (30). The A-utterance was repeated from the sentence from the previous page. The B-response offered a confirmation of A's proposition content. The participant was next given a binary forced choice question:

(30) A: Dajia dou bu gen xiaoming shuohua. Guji ta shenme shihou
everyone DOU NEG with Xiaoming talk by.estimate he what time
dezui-guo ren.
offend-EXP person
'No one's talking to Xiao Ming. Guess he rubbed the folks up the wrong way at some point.'

B: Haoxiang shi.
apparently COP
'It appears so.'

Question: Is the response by B to the statement by A natural to you?

A. Yes B. No

The reply of B makes for a felicitous continuation in the ongoing discourse, under the assumption that the target sentence of A is uttered as a non-interrogative/declarative statement. B's reply will not be seen as felicitous, however, with an interrogative construal of the A-utterance: By being cooperative, B is expected to provide new information relative to the wh-portion, not confirmation. In other words, if a non-interrogative reading is what participants consistently obtained, we should expect that they overwhelmingly picked the *Yes*-choice. We could thus reaffirm whether participants successfully obtained the intended interpretation by inspecting the percentage of *Yes*-choices for the target items. This was evaluated against the baselines of filler sentences, which were interspersed with the target stimuli items at a 2:1 ratio in the design. Half of the fillers received a canonically interrogative reading (e.g. 31a), for which we expected that participants overwhelmingly picked the *No*-choice. The other half were declarative sentences (e.g. 31b), for which we expect predominantly *Yes*-choices.

(31) a. Haizi-men zai na'er canjia xialingying huodong?
child-PL LOC where take.part summer.camp activities
'Where do the kids take part in summer camp activities?'

b. Daxiongmao zai caodi shang chi-zhe zhuzi.
giant.panda LOC where like live-PRF
'The giant panda was eating bamboo on the lawn.'

Inspecting the forced choice percentages for both filler types thus offers us a baseline for comparison with the percentage of choices in the target items.

Target sentences followed a Latin square design. 18 target sentences were divided into three lists, with participants rating the target stimuli assigned to three groups accordingly. Each individual rated 6 target sentences and 12 filler sentences. Control items followed the same Latin square arrangement, yielding another three lists. In total, 720 data points entered the statistical analysis (360 for each subexperiment).

The procedures of Experiment 3 were similar to those in Experiment 1 and 2, with the exception that each page containing the stimulus items was followed by another page containing a corresponding forced-choice comprehension question. The stimulus items were pseudo-randomly interspersed with twice as many fillers serving as distractors. The entire experiment, including the background survey, lasted around 8 minutes.

Results As in the previous experiments, **Table 3** reports the key point estimates, separately presented by subexperiment. **Figure 3** shows the ratings by condition for the first (wh-indefinites) subexperiment after z-score transformation.

Figure 4 shows the ratings by condition for the second (wh-free choice) subexperiment after z-score transformation.

Our visual inspection showed that across the three wh-expressions, the 'why' condition received the lowest rating, regardless of whether it occurs in the indefinite construction or the free-choice construction. In comparison, the 'what time' condition and the 'how' condition were rated as generally acceptable in both contexts, with 'what time'-sentences receiving the most natural judgment by the participants. For the control conditions, ratings were generally high across all three wh-categories, with minimal differences between them. Mean z-scores of control items were similar to those of target wh-nominal and target 'how' sentences, but not to target why, indicating that controls patterned with the higher-rated target items. The scatterplots also showed that participant ratings for controls were more tightly clustered than those for targets, reflecting greater uniformity in acceptability judgments.

This pattern was further confirmed by two ordinal mixed models applied to the indefinite and free choice environment separately. Cumulative link mixed models (`clmm(Rating~ wh.category * type + (1|participant)+ (1|item)`) were fitted separately to the two sub-experiments, each applied to its respective data set.

Table 3: Z-scored ratings in Experiment 3.

Condition type		Wh-type	z-scored means (SD)	95% CI
Indefinites	Target	whnom	0.78 (0.58)	[0.54, 1.02]
	Target	how	0.70 (0.47)	[0.46, 0.94]
	Target	why	−0.41 (0.67)	[−0.65, −0.16]
	Control	whnom	0.77 (0.50)	[0.62, 0.92]
	Control	how	0.91 (0.27)	[0.76, 1.06]
	Control	why	0.69 (0.49)	[0.54, 0.84]
Free choice	Target	whnom	0.66 (0.52)	[0.40, 0.92]
	Target	how	0.61 (0.91)	[0.34, 0.87]
	Target	why	−0.42 (0.53)	[−0.69, −0.16]
	Control	whnom	0.64 (0.53)	[0.44, 0.85]
	Control	how	0.56 (0.48)	[0.36, 0.77]
	Control	why	0.62 (0.57)	[0.42, 0.83]

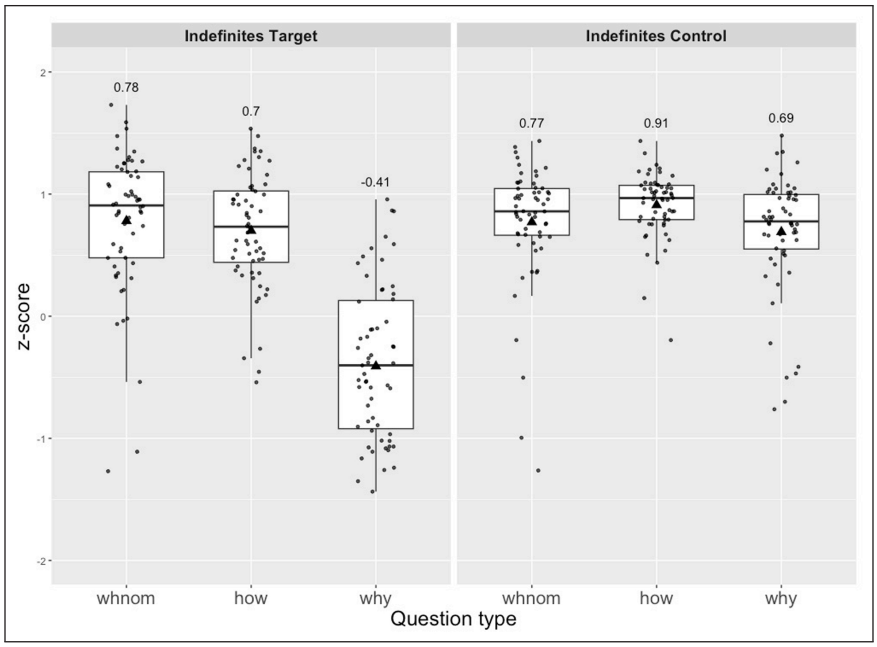

Figure 3: A box plot, augmented with a scatterplot, showing the z-score transformed ratings across wh-type pertaining to the indefinites paradigm (left panel: target; right panel: control).

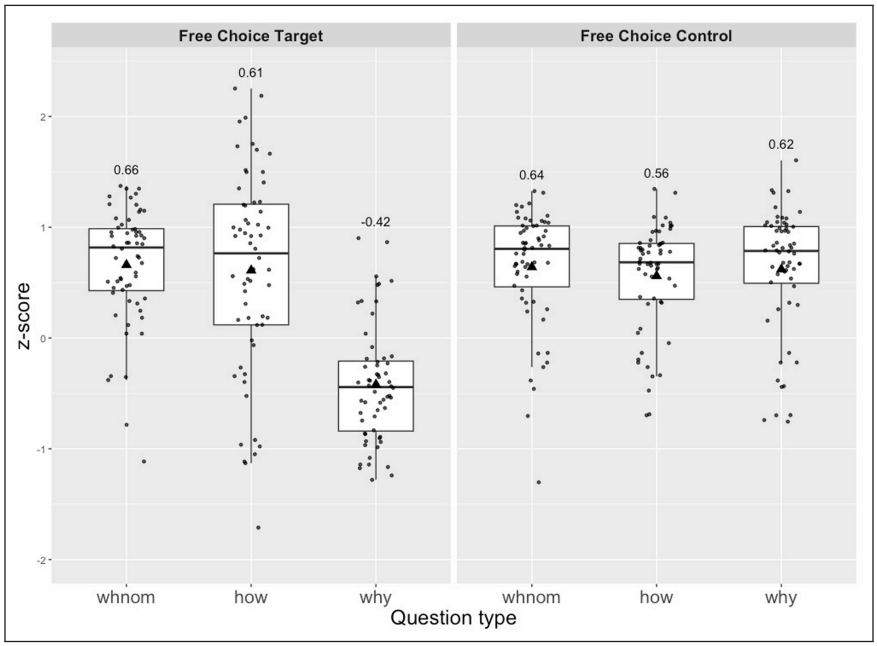

Figure 4: A box plot, augmented with a scatterplot, showing the z-score transformed ratings across wh-type pertaining to the free choice paradigm (left panel: target; right panel: control).

The models predicted the ratings from a fixed effect of `wh.category` (contrast coded against the 'how'-condition) interacting with `type` (`target`, `control`), as well as by-`participant` and by-`item` random effects. For wh-indefinites, our models revealed a significant interaction between `wh.category` and `type`, indicating that the effect of wh-category differed between target and control items. Separately testing the individual pairwise comparisons within the target and the control conditions using post hoc HSD tests (with α-adjustments), our results revealed that compared with 'how'-indefinites, a higher rating for the wh-nominal indefinites condition ($\hat{\beta}$: 0.83±0.39, $p \approx 0.08$) and a significantly lower rating for the why-indefinites condition ($\hat{\beta}$: −3.93±0.53, $p < 0.0001$) were observed. There was no significant difference between each pair of the three control conditions. For the free choice environment, the model also revealed a significant interaction between `wh.category` and `type`. Within the target conditions, post hoc comparisons showed that the 'why' condition is rated significantly lower than the 'how' condition ($\hat{\beta}$: −2.11±0.423, $p < 0.001$), as well as the 'what time' condition ($\hat{\beta}$: −3.03±0.48, $p < 0.001$). There was no significant difference between the 'what time' condition and the 'how' condition. In addition, there was no significant difference between each pair of the three control conditions.

Overall, our results show that 'why' was rated the least acceptable when interpreted as an indefinite and as a free choice item. Between 'what time' and 'how' conditions, a

higher rating was observed for 'what time' across both environments, which did not reach statistical significance.

Our additional inspection of forced choice results revealed a 95% rate of *Yes*-choices for 'what time' and 'how'-sentences in the wh-indefinite and wh-free choice contexts, i.e. all the acceptable target sentences. For a comparison, the 'good' filler sentences (i.e. judged to be acceptable by participants along the Likert-scale) registered a mean accuracy rate of 89% (averaged over declarative and interrogative sentence types, we considered a forced choice response to declaratives to be accurate if it yielded a *Yes*-choice, and one to interrogatives to be accurate if it yielded a *No*-choice). We take the fact that participants overwhelmingly chose 'Yes'-responses to indicate unequivocally that participants in general are capable of incorporating the lexical, grammatical and orthographical information in the immediate context to obtain a non-interrogative reading for the target sentences, at least for all the acceptable items in our experiment.

4 Discussion

The current study presented three experiments to examine how Mandarin speakers rated sentences with causal 'why'. Our rating paradigms compared the Likert-scale-based acceptability judgments of why-sentences with those of two other wh-categories, i.e. *zenme* 'how' and *shenme shihou* 'what time'. Experiment 1 tested the effect of wh-category on sentence acceptability under a focus-intervention configuration. Experiment 2 looked specifically into the between-category difference within a multiple wh-interrogative environment. Experiment 3 investigated whether wh-category had an influence on acceptability under the context that licenses an indefinite as well as free choice reading.

We offer several findings. First, results from Experiment 1 revealed that all three wh-expressions incurred a degraded judgment when c-commanded by focus, in contrast to the control conditions, which were consistently rated above the midrange on the z-transformed scale. Between each target condition and its corresponding control, we observed a significant difference. Still, when we only take account of target sentences, why-sentences received significantly more degraded judgments than those with both 'how' and 'what time'. The acceptability of 'how' and 'what time' were similarly degraded under focus-induced intervention, for which our model found no significant difference. Hence the evidence pointed to a two-way difference ('why' vs. others) within intervention-violating configurations.

Second, results from Experiment 2 showed that why-sentences were unacceptable in multiple wh-in-situ contexts. We did not observe a categorial distinction between 'how' and 'what time': Regardless of whether the phrase was nominal or not, judgments tended towards being

acceptable. We conclude that a two-way distinction between 'why' and other wh-types best accounted for the pattern of judgments within the environment of multiple wh-questions.

An anonymous reviewer observed that the acceptability of 'why' in multiple-wh configurations could be order-sensitive. Specifically, configurations where the controlled wh-element precedes 'why' are more acceptable than those where it follows 'why'. We wish to emphasize that in our design, we consistently used the configuration that the reviewer identifies as having a potential ameliorating effect: the target *weishenme* was preceded by the second wh-element, *shui* 'who' (cf. example 24a). Even with this preferred order, we found a significant effect of wh-type. In other words, under the potentially rescuing 'who-before-why' order, the 'why' condition received distinctively low ratings compared to other wh-categories. This supports our claim regarding the special behavior of 'why'. If there is a *minimal-pair* difference regarding order, we would expect the 'why-before-who' order to yield even worse ratings than those reported here.

Tentatively, the observation regarding the potential amelioration with the second wh-element preceding 'why' suggests that multiple-wh configurations may interact with exceptional wide scope mechanisms in the formal semantics literature (e.g. a scope-out/quantify-in analysis, cf. Krifka 2001; Kitagawa et al. 2003; Fox 2012). While a detailed discussion is beyond the scope of the current, empirically-oriented study, this remains an interesting field to explore to better delineate the special behavior of *weishenme*.

Third, in Experiment 3, as with the previous two experiments, our results revealed an effect of wh-category on the acceptability judgment of target sentences, which were consistent across the two non-question environments (indefinite and free-choice) investigated. Among them, why-sentences were consistently rated low, with a significant difference observed against the ratings of 'how'-sentences as well as 'what time'-sentences. Our results did not yield a finding of significance with regards the comparison of 'how' and 'what time'. Both types elicited generally acceptable judgments. We conclude that 'why' differs from other wh-types in being excluded from non-question readings.

In sum, the results have converged on a significant finding regarding the why-adjunct across the three experimental settings. Importantly, these findings cannot be explained away by invoking a broader class defined on nominality. This is because we controlled our experiments to include 'how' as a baseline of comparison. If a distinction along the line of nominality underlies the distinctive behaviors observed for the why-adjunct, we would expect that 'why' and 'how' behave on par with one another, which did not turn out to be the case: What was revealed was a robust difference of clear effect size between 'why' and 'how' that were consistent across experiments. In each one of them, 'why' induced the more robust violations: 'Why' incurred stronger degradation in judgment than 'how' in intervention-violating environments. In both multiple-wh and non-interrogative environments, a significant decrease in acceptability

compared against 'what time' was obtained for 'why', but not for 'how'. Taken together, we are able to establish that 'why' (instead of *generic* adjuncts) is uniquely incapable of yielding a coherent interpretation.

As section 2 detailed, establishing these empirical patterns constitutes evidence for the wide scope view of the why-adjunct (Ko 2005; 2006; Tsai 2008; Stepanov & Tsai 2008; Tomioka 2009; Takita & Yang 2014; Endo 2015; Jin 2016; Miyagawa 2017; Murphy 2017; Cheng 2021). According to this view, 'why' does not leave a trace. Rather, it is directly base-generated where it takes scope, which is assumed to be the scope position of the Q operator (Lawler 1971; Bromberger 1992; Tomioka 2009; Stepanov & Tsai 2008; Soare 2021). It thus resides in the C domain instead of the T domain, unlike the other wh-adjuncts. Assuming this, the intervention configuration where focus c-commands 'why' is then not a possible one, as the focus operator must stay within the domain of the Q operator. Next, in a multiple-wh configuration, with 'why' entering compositional interpretation from a structure above the C head, it requires a propositional argument specified with a true value, which conflicts with the necessarily unspecified nature of the question radical it ends up composing with. Finally, in both existential and universal wh-readings, the high attachment position of 'why' has the consequence that its Hamblin alternatives cannot be evaluated by an existential or universal operator. As such, the body of experimental evidence lends support to a growing body of theoretical research regarding the scopal behaviors of 'why' (Lawler 1971; Cattell 1978; Bromberger 1992; Hegarty 1992; Iatridou & Kroch 1992; Oshima 2007; Tomioka 2009; Soare 2017; 2021).[14] A caveat is in order, however, as the current experimental evidence directly motivates the case for a wide scope analysis of Mandarin *weishenme*, and likely carries over to its counterparts in other East Asian wh-in-situ languages. It is, in principle, possible that other languages behave differently. In particular, a potential point of variation lies with languages in which wh-elements regularly undergo movement to take scope. For instance, it has been argued that in at least a subset of these languages, the why-adjunct undergoes auxiliary inversion, which is traditionally taken as a diagnostic for movement (Rizzi 2001; Ko 2005; Thornton 2008; Shlonsky & Soare 2011). It is then possible that the why-element in such languages moves to take scope after all, just like the other wh-elements. We therefore leave open the possibility of parametrization, in the sense that the why-element across languages does not uniformly instantiate the same "no-trace, direct-merge" option. Understanding how this parametrization pans out remains an intriguing next step.[15]

In addition to the empirical support for the high attachment of 'why', our experiments yield additional findings. First, we observed a pattern of degraded judgment for the non-nominal

[14] See independent work from island domains (Jin & Yan 2025; Ji & Jin 2025) that converge on the wide scope behaviors of 'why'.

[15] Note that wh-movement languages are not homogeneous with regards inversion: Italian *perché* 'why' is observed to not trigger auxiliary inversion.

wh-adjunct 'how' as well as the nominal 'what time' in an intervening environment. Second, a mild degradation of 'how' was found in the multiple-wh context. We attempt preliminary explanations of these findings and discuss the implications. We will leave it to future experiments to validate these explanations.

Experiment 1 additionally revealed that 'how' and 'what time' elicited degraded judgments when c-commanded by a focus phrase. While the size of degradation was smaller compared to that of 'why', it was still found that participants rated the intervention-violating structures of the two wh-types lower than the control structures and also the middling range of the scale of judgment, which we consider as evidence for a canonical case of focus intervention effect. Our findings more specifically show a uniform, consistent decline in acceptability across nominality lines. In this aspect the findings replicate the experimental results in Jin & Yan (2024) and confirm a widely held view regarding the particular robustness of focus as an intervener (Kim 2002; 2006; Beck 2006; Soh 2005; Hagstrom 2006; Yang 2009; 2011). A key assumption has been that focus triggers intervention regardless of the wh-category, even if a wh-type distinction might be relevant for other scopal elements (like quantifiers) (Tsai 1994; Cheng & Rooryck 2000; Law 2001; Soh 2005). By confirming this assumption, our findings favor an explanation of focus intervention based on the notion that focus and wh are interpreted in much the same way. Wh-phrases range over Hamblin alternatives (Karttunen 1977), just like focus triggers context-determined alternatives (Rooth 1995). Both the Q operator and the focus operator are sensitive to the alternatives within their domain. Such parallelism has the consequence that when an offending focus operator is closer to wh than the Q operator is, the evaluation of wh-alternatives by Q becomes affected of sorts (Beck 2006; Cable 2010; Li & Law 2016). Regardless of the exact theoretical implementation, the idea pertains to the fundamental alternative semantics of wh, hence is independent of wh-type. This way, focus evaluation accounts are insensitive to the categorial difference between wh-types.[16]

An interesting further question has to do with capturing the relatively milder pattern of unacceptability for the non-'why', garden-variety wh-intervention effect. Some accounts argue that the underlying explanation lies in pragmatics or working memory. A pragmatic explanation to wh-intervention (e.g. Tomioka 2007; Eilam 2008; Branan 2017) builds on the assumption that linearly adjacent positions under a canonical information structural configuration are limited to a partition of new information and old information. Focus encodes new information, and in a question-answer sequence, the wh-element corresponds to the focused constituent in the answer. The simultaneous occurrence of a focus phrase and a wh-expression thus creates an anomalous case to the normal way information is packaged. Alternatively, it is proposed that

[16] The empirical picture is, however, more complicated than presented here. Xie (2014) argues that non-exhaustive focus does not trigger intervention effects. It is also proposed that focus adverbs in Mandarin do not act as interveners, unlike focus DPs (Soh 2005; Yang 2009; 2011). We will leave it to future work to tackle different focus types.

parsing a focus expression is a process of dependency resolution (of the association between focus and its operator). With wh-expression interpreted akin to focus, the intervention-violating configuration leads to a nested dependency of multiple foci, which is independently known to be difficult to process (Lewis & Vasishth 2005; Vasishth & Lewis 2006; Beck & Vasishth 2009; Jin & Yan 2024). In each of these two cases, we expect to see more individual differences in the ability to accommodate pragmatic violations or in the cognitive capacity relevant to processing, which results in more variation in judgment. We expect to explore this issue further in the future.

In Experiment 2, 'how' elicited a discernible decline in judgments compared to 'what time' in multiple-wh contexts. Note that this decline did not reach statistical significance, with 'how'-sentences in general rated above the middling range on the z-transformed scale of judgment. Still, we attempt some explanation here. We suspect d(iscourse)-linking plays a role, in line with a vast literature independently attributing differences in acceptability judgment across wh-category to the variation in d-linking status. Here we understand d-linking as a notion lying at the syntax-discourse interface, as outlined in Comorovski (1987) and Pesetsky (1987). A d-linked 'wh'-phrase is, semantically speaking, one that requires the range of felicitous answers to a question be limited to a contextually salient set. As it stands, phrases denoting individuals (including places and times) more easily settle upon a restricted domain than phrases denoting manners. It is argued that acceptability judgments correlate with the extent to which a phrase lends itself to domain restriction (Anagnostopoulou 1994). The fact that 'what time' invites better judgments than 'how' could be because the former introduces a lexical restriction and settles upon a salient domain from which the alternatives are chosen, and the latter is less readily so.

In addition, the d-linking factor may exert a stronger drag on acceptability in multiple-wh contexts than in single-wh ones. On a related note, we observed that in forming a multiple wh-question that solicits information about the manner of action, wh-expressions with an explicit lexical restriction were preferred over *zenme* 'how'. In an informal follow-up consultation with three native speakers not involved in the main experiments, we asked how they would form a multiple 'wh'-question eliciting both an individual person and a manner, based on materials taken from our stimuli from Experiment 2. In these cases, the favored strategy made use of alternative forms as *yong shenme fangshi* 'with what manner' or *yong shenme shouduan* 'with what means', e.g. *Shui yong shenme fangshi bangjia-le fushang Huo xiansheng?* 'Who kidnapped Mr. Huo the business tycoon in what manner?'. While the use of *zenme* 'how' was still more often than not comprehensible, it was not considered the most common way to pose the relevant question in actual speech. This raises the possibility that the infelicity of *zenme* 'how' is Gricean by nature, as the (more ready) availability of alternative forms affects how speakers accept *zenme* in questions. Future comparison with existing studies on the amelioration effect of d-linking or the effect of constructing alternatives in pragmatic reasoning on wh-questions (e.g.

Kluender 1998; Hofmeister & Sag 2010; Goodall 2015) are needed to understand how broadly these pragmatic factors may apply (cf. Chen 2023 for work on Mandarin).

5 Conclusions

This paper presents results from controlled experiments showing that the why-adjunct in Mandarin behaves in a unique way. A total of three formal judgment surveys we conducted in intervention-violating contexts, multiple wh-interrogative contexts, as well as non-interrogative wh-contexts, all point in the same direction: 'why' presents an exception to the generalization that characterizes wh-questions. This result holds when a fine-grained distinction within wh-phrases is considered. We conclude that the evidence is compatible with a high attachment approach to 'why', argued for in Rizzi (1990), Bromberger (1992), inter alia, according to which 'why' does not leave a trace during interpretation as regular wh-expressions do and is restricted to the widest scope.

Appendix

The list of abbreviations in this paper is given as follows.

ACC	accusative	IPFV	imperfective	Q	question particle
CLF	classifier	LOC	locative	REL	relativizer
COP	copula	NEG	negation	RES	resultative
DAT	dative	NOM	nominative	RP	resumptive pronoun
DEM	demonstrative	PL	plural	SG	singular
DOU	universalizing adverb/particle	POSS	possessive	TOP	topic marker
ERG	ergative	PRF	perfect		
EXP	experiential aspect	PRS	present		
FUT	future tense	PRT	particle		
GEN	genitive	PST	past		
HAB	habitual	PTCP	participle		

Table A1: Abbreviations of glossing terms.

Data Availability Statement: The stimuli lists used in the three experiments of our survey research, as well as a description of the complete statistical analysis and the R code scripts, are stored privately in the OSF repository. We plan to make them accessible in the event that the manuscript is accepted for publication.

Competing Interests: The authors have no competing interests to declare.

Funding: This study was funded by The Philosophy and Social Sciences Office of the Government of Shanghai (grant No. 2021BYY007).

Acknowledgements: We wish to express our deep gratitude to the two anonymous reviewers of *Zeitschrift für Sprachwissenschaft* and the handling editor, Professor Oliver Schallert, for their careful reading and detailed comments, which have improved the quality of the manuscript. We further thank Hanbo Yan for her very valuable help on methodology and data analysis. Early stages of this research benefitted directly from discussions with Xingbao Niu and Ali Hussain Birahimani. The manuscript also received valuable feedback and inspiration through discussions with Sigrid Beck, Aron Hirsch, Tatiana Scheffler, Swantje Tönnis, and Matthias Gerner. Xuziyun Pan and Yan Yang kindly assisted with collecting data and recruiting participants. Lihua Xu, Rick Yan, Min Li, and Yu Li offered support at various stages of this project. We are indebted to the above individuals. Finally, we appreciate the audiences at the linguistic meetings at GDUFS, Aarhus University and Wuhan University. As usual, all the errors in this paper are our own.

References

Anagnostopoulou, Elena. 1994. *Clitic dependencies in Modern Greek*. University of Salzburg dissertation.

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14(1). 1–56. https://doi.org/10.1007/s11050-005-4532-y.

Beck, Sigrid. 2019. Which alternatives? Talk given at the DgfS XPrag summer school.

Beck, Sigrid & Shravan Vasishth. 2009. Multiple focus. *Journal of Semantics* 26. 159–184. https://doi.org/10.1093/jos/ffp001.

Belnap, Nuel. 1969. Questions: Their presuppositions, and how they can fail to arise. In Kirk Lambert (ed.), *The logical way of doing things*, 23–37. New Haven: Yale University Press.

Branan, Kenyon. 2017. In-situ wh-phrases in superiority violating contexts don't have to be in-situ. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *A pesky set: Papers for David Pesetsky*, 353–359. Cambridge, MA: MITWPL.

Bromberger, Sylvain. 1992. *On what we know we don't know: Explanation, theory, linguistics, and how questions shape them*. Chicago: University of Chicago Press.

Bruening, Benjamin. 2007. Wh-in-situ does not correlate with wh-indefinites or question particles. *Linguistic Inquiry* 38(1). 139–166. https://doi.org/10.1162/ling.2007.38.1.139.

Cable, Seth. 2010. *The grammar of Q*. Oxford: Oxford University Press.

Cattell, Ray. 1978. On the source of interrogative adverbs. *Language* 54(1). 61–77. https://doi.org/10.2307/412999.

Cecchetto, Carlo & Caterina Donati. 2017. Why there is no non-interrogative *why*. Paper presented at 2017 SLE Annual Meeting, Zurich.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Jun. 2023. *On shi and de in Mandarin: clefts and beyond*. Universität Stuttgart dissertation.

Cheng, Lisa. 1991. *On the typology of wh-questions*. MIT dissertation.
Cheng, Lisa. 2021. What-as-why sentences in Cantonese. In Gabriela Soare (ed.), *Why is why unique?*, 219–248. Berlin: de Gruyter Mouton.
Cheng, Lisa & Johan Rooryck. 2000. Licensing wh-in-situ. *Syntax* 3(1). 1–19.
Christensen, Rune Haubo Bojesen. 2019. Ordinal—-regression models for ordinal data. Manuscript, Technical University of Denmark.
Comorovski, Illeana. 1987. Discourse-linking and the wh-island constraint. *Papers of the North East Linguistics Society* 19(1). 254–274.
Dayal, Veneeta. 2017. *Questions*. Oxford: Oxford University Press.
Eilam, Aviad. 2008. Intervention effects: Why Amharic patterns differently. In Natasha Abner & Jason Bishop (eds.), *WCCFL 27*, 141–149. Somerville, MA: Cascadilla Proceedings Project.
Endo, Yoshio. 2015. Two ReasonPs: What are*(n't) you coming to the US for? In Ur Shlonsky (ed.), *Beyond functional sequence*, 220–231. New York: Oxford University Press.
Engdahl, Elisabet. 1986. *Constituent questions*. Dordrecht: Reidel.
Ernst, Thomas. 1994. Conditions on Chinese A-not-A questions. *Journal of East Asian Linguistics* 3(3). 241–264. https://doi.org/10.1007/BF01733065.
Ernst, Thomas. 2001. *The syntax of adjuncts*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/cbo9780511486258.
Fortmann, Christian. 2019. *Warum*-infinitives in German. *Linguistische Berichte* 258. 179–227. https://doi.org/10.46771/2366077500258_2.
Fox, Danny. 2012. Multiple wh-questions: uniqueness, pair-list and second order questions. MIT Seminar.
Giannakidou, Anastasia & Lisa Cheng. 2006. (In)definiteness, polarity and the role of wh-morphology in free choice. *Journal of Semantics* 23(2). 135–183. https://doi.org/10.1093/jos/ffl001.
Gibson, Ted & Evelina Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14(6). 233–234. https://doi.org/10.1016/j.tics.2010.03.005.
Gibson, Ted, Steven Piantadosi & Evelina Fedorenko. 2013. Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes* 28(3). 229–240. https://doi.org/10.1080/01690965.2012.70.
Goodall, Grant. 2015. The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology* 5. 1493. https://doi.org/10.3389/fpsyg.2014.01493.
Hagstrom, Paul. 2006. A-not-A Questions. In Henk van Riemsdijk & Martin Everaert (eds.), *The Blackwell companion to syntax*, 173–213. London: Blackwell.
Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Clarendon Press.
Hegarty, Michael. 1992. Adjunct extraction without traces. In *Proceedings of the 10th West Coast Conference on Formal Linguistics*, 209–223.
Hintikka, Jaakko & Ilpo Halonen. 1995. Semantics and pragmatics for *why*-questions. *Journal of Philosophy* 92(12). 636–657. https://doi.org/10.2307/3808201.
Hofmeister, Philip & Ivan Sag. 2010. Cognitive constraints and island effects. *Language* 86(2). 366–415.
Hoji, Hajime. 1985. *Logical form constraints and configurational structures in Japanese*. University of Washington doctoral dissertation.
Hu, Fang. 2002. A prosodic analysis of wh-words in Standard Chinese. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, 11–13 April.

Huang, James. 1982. *Logical relations in Chinese and the theory of grammar*. MIT dissertation.
Huang, James, Audrey Li & Yafei Li. 2009. *The syntax of Chinese*. Cambridge, UK: Cambridge University Press.
Iatridou, Sabine & Anthony Kroch. 1992. The licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon. *Working Papers in Scandinavian Linguistics* 50. 498–507.
Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlỳ & Vít Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference* , 125–127. Valladolid.
Ji, Xu & Dawei Jin. 2025. Comparing *wh-in-situ* and *wh*-fronting island effects. Manuscript, Shanghai Jiao Tong University.
Jin, Dawei. 2016. *The semantics-pragmatics interface and island constraints in Chinese*. University at Buffalo dissertation.
Jin, Dawei. 2019. A semantic account of intervention effects in Chinese *why*-questions. *Linguistics and Philosophy* 52(1). 1–43. https://doi.org/10.1007/s10988-019-09270-x.
Jin, Dawei. 2020. In situ islands are root phenomena. Manuscript, Shanghai Jiao Tong University. https://doi.org/10.13140/RG.2.2.26005.14567/1.
Jin, Dawei. 2022. The intervention effect in Suzhounese questions. *Poznań Studies in Contemporary Linguistics* 58(4). 775–822. https://doi.org/10.1515/psicl-2022-0031.
Jin, Dawei & Hanbo Yan. 2024. Intervention effects in Mandarin Chinese—An experimental study. *Journal of Semantics* 41(2). 121–148. https://doi.org/10.1093/jos/ffae006.
Jin, Dawei & Hanbo Yan. 2025. Reassessing the argument-adjunct asymmetry in wh-in-situ islands in Mandarin: An experimental investigation. *Natural Language and Linguistic Theory* 43. 1327–1363. https://doi.org/10.1007/s11049-024-09638-0.
Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 3–44.
Kim, Nayeon, Ziying Li & Jiayi Lu. 2023. Island sensitivity of two different interpretations of why in Chinese. *Frontiers in Psychology* 13. 1059823. https://doi.org/10.3389/fpsyg.2022.1059823.
Kim, Shin-Shook. 2002. Focus matters: Two types of intervention effect. Talk presented at WCCFL 21, Santa Cruz.
Kim, Shin-Shook. 2006. *Intervention effects in questions*. University of Frankfurt doctoral dissertation.
Kitagawa, Yoshihisa, Dorian Roehrs & Satoshi Tomioka. 2003. Multiple *wh*-interpretations. In *Generative Grammar in a broader perspective: Proceedings of the 4th GLOW in Asia*, 209–233.
Kluender, Robert. 1998. The limits of syntax. In *On the distinction between strong and weak islands: A processing perspective*, 241–279. Amsterdam: Brill.
Ko, Heejeong. 2005. Syntax of why-in-situ: Merge into [Spec, CP]. *Natural Language & Linguistic Theory* 23(4). 867–916. https://doi.org/10.1007/s11049-004-5923-3.
Ko, Heejeong. 2006. On the structural height of reason wh-adverbials: Acquisition and consequences. In Lisa Cheng & Norbert Corver (eds.), *Wh-movement moving on*, 319–349. Cambridge, MA: MIT Press.
Kotek, Hadas. 2014. *Composing questions*. MIT dissertation.
Kotek, Hadas. 2017. Intervention effects arise from scope-taking across alternatives. *NELS* 47. 1–14.
Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Chung-min Lee, Ferenc Kiefer & Manfred Krifka (eds.), *Contrastiveness in information structure, alternatives and scalar implicatures*, 123–143. Berlin: Springer.

Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9(1). 1–40. https://doi.org/10.1023/A:1017903702063.

Kuroda, Shigeyuki. 1965. *Generative grammatical studies in the Japanese language*. MIT dissertation.

Larson, Richard & Miyuki Sawada. 2012. Root transformations and quantificational structure. In Liliane Haegeman, Nye Richard & Aelbrecht Lobke (eds.), *Main clause phenomena: New horizons*, 47–77. Amsterdam: John Benjamins.

Law, Ann. 2001. A-not-A questions in Cantonese. *UCL Working Paper in Linguistics* 13. 295–318.

Lawler, John. 1971. Any questions? In *Chicago Linguistic Society 7*, 163–173.

Lewis, Richard & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29. 1–45. https://doi.org/10.1207/s15516709cog0000_25.

Li, Audrey. 1992. Indefinite wh in mandarin chinese. *Journal of East Asian Linguistics* 1(2). 125–155. https://doi.org/10.1007/BF00130234.

Li, Haoze. 2011. *Focus intervention effects in Mandarin*: Chinese University of Hong Kong MA thesis.

Li, Haoze & Jess Law. 2016. Alternatives in different dimensions: a case study of focus intervention. *Linguistics and Philosophy* 39(3). 201–245. https://doi.org/10.1007/s10988-016-9189-0.

Lin, Jo-Wang. 1998. On existential polarity wh-phrases in chinese. *Journal of East Asian Linguistics* 7(3). 219–255. https://doi.org/10.1023/A:1008284513325.

Lin, Jo-Wang. 2004. Choice functions and scope of existential polarity wh-phrase in Mandarin Chinese. *Linguistics and Philosophy* 27(4). 451–491. https://doi.org/10.1023/B:LING.0000024407.76999.f7.

Linzen, Tal & Yohei Oseki. 2018. The reliability of acceptability judgments across languages. *Glossa* 3(1). 1–25. https://doi.org/10.5334/gjgl.528.

Lü, Shuxiang. 1942. *Zhongguo Wenfa Yaolue [A synopsis of Chinese grammar]*. Beijing: The Commercial Press.

Mellor, D.H. 1995. *The facts of causation*. London: Routledge.

Mellor, D.H. 2004. For facts as causes and effects. In J. Collins, N. Hall & L. Paul (eds.), *Causation and counterfactuals*, 309–324. Cambridge, MA: MIT Press.

Miyagawa, Shigeru. 2017. *Agreement beyond phi*. Cambridge: MIT Press.

Murphy, Andrew. 2017. Toward a unified theory of wh-in-situ and islands. *Journal of East Asian Linguistics* 26(1). 189–231. https://doi.org/10.1007/s10831-017-9155-z.

Niu, Xingbao. 2023. *A comparative study of the syntax and semantics of causal interrogative questions in Chinese and Korean*: Shanghai Jiao Tong University MA thesis.

Oshima, David. 2007. On factive islands: Pragmatic anomaly vs. pragmatic infelicity. In Takashi Washio, Ken Satoh, Hideaki Takeda & Akihiro Inokuchi (eds.), *New frontiers in artificial intelligence*, 147–161. Berlin: Springer. https://doi.org/10.1007/978-3-540-69902-6_14.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice ter Meulen (eds.), *The representation of (in) definiteness*, 98–129. Cambridge, MA: The MIT Press.

Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20(4). 335–397.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: The MIT Press.

Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. *Current studies in Italian syntax* 14. 267–296.

Rooth, Mats. 1995. *Association with focus*. Amherst, MA: University of Massachusetts dissertation.

Schwarz, Bernhard & Alexandra Simolenko. 2018. On the logical makeup of how- and why-questions. In Sireemas Maspong et al. (ed.), *Proceedings of semantics and linguistic theory 28*, 533–545. Washington DC: Linguistic Society of America.

Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Berlin: Language Science Press. https://doi.org/10.17169/langsci.b89.100.

Schütze, Carson T. & Jon Sprouse. 2014. Judgment data. In Robert Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge, UK: Cambridge University Press.

Shi, Hongli. 2021. *A study on the syntax and semantics of non-interrogative question words in Chinese*. Beijing Foreign Studies University dissertation.

Shlonsky, Ur & Gabriela Soare. 2011. Where's 'why'? *Linguistic Inquiry* 42(4). 651–669. https://doi.org/10.1162/LING_a_00064.

Soare, Gabriela. 2017. Why is 'why' Unique? Talk given at the workshop at the 2017 annual SLE meeting.

Soare, Gabriela. 2021. *Why is 'why' Unique?: Its syntactic and semantic properties*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110675160.

Soh, Hooi Ling. 2005. Wh-in-situ in Mandarin Chinese. *Linguistic Inquiry* 36(1). 143–155. https://doi.org/10.1162/ling.2005.36.1.143.

Sprouse, Jon, Shin Fukuda, Hajime Ono & Robert Kluender. 2011. Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax* 14(2). 179–203. https://doi.org/10.1111/j.1467-9612.2011.00153.x.

Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from linguistic inquiry. *Lingua* 134. 219–248. https://doi.org/10.1016/j.lingua.2013.07.002.

Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. https://doi.org/10.1353/lan.2012.0004.

Stepanov, Arthur & Wei-Tien Dylan Tsai. 2008. Cartography and licensing of *wh*-adjuncts: a cross-linguistic perspective. *Natural Language & Linguistic Theory* 26(3). 589–638. https://doi.org/10.1007/s11049-008-9047-z.

Sugimura, Hirofumi. 2007. Xiandai Hanyu Yiwen Daici Ye/dou Jiegou de Yuyi Fenxi [A semantic explanation of the universal use of modern Chinese interrogative pronouns]. *Shijie Hanyu Jiaoxue* 17(3). 284–301.

Szabolcsi, Anna & Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1. 235–284. https://doi.org/10.1007/BF00263545.

Takita, Kensuke & Barry Chung-Yu Yang. 2014. On multiple wh-questions with 'why' in Japanese and Chinese. In Mamoru Saito (ed.), *Japanese syntax in comparative perspective*, 206–227. New York: Oxford University Press.

Thornton, Rosalind. 2008. Why continuity. *Natural Language & Linguistic Theory* 26(1). 107–146. https://doi.org/10.1007/s11049-007-9031-z.

Tomioka, Satoshi. 2007. Pragmatics of LF intervention effects. *Journal of Pragmatics* 39(9). 1570–1590. https://doi.org/10.1016/j.pragma.2007.03.002.

Tomioka, Satoshi. 2009. Why questions, presuppositions, and intervention effects. *Journal of East Asian Linguistics* 18(4). 253–271. https://doi.org/10.1007/s10831-009-9053-0.

Tsai, Wei-Tien Dylan. 1994. On nominal islands and LF extraction in Chinese. *Natural Language & Linguistic Theory* 12(1). 121–175. https://doi.org/10.1007/BF00992747.

Tsai, Wei-Tien Dylan. 1999. The *how*s of *why* and the *why*s of *how*. *UCI Working Papers in Linguistics* 5. 155–184.

Tsai, Wei-Tien Dylan. 2008. Left periphery and how-why alternations. *Journal of East Asian Linguistics* 17(2). 83–115. https://doi.org/10.1007/s10831-008-9021-0.

Vasishth, Shravan & Richard Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82. 767–794.

Xie, Zhiguo. 2014. Focus, (non-)exhaustivity, and intervention effects in wh-in-situ argument questions. *The Linguistic Review* 30(4). 585–617. https://doi.org/10.1515/tlr-2013-0019.

Xue, Xiaoying. 2014. An investigation into the non-interrogative uses of question words in Chinese. *Journal of Hunan University of Science and Technology* 17(3). 128–132.

Yang, Barry Chung-Yu. 2009. *Intervention effects and the covert component of grammar*. National Tsinghua University dissertation.

Yang, Barry Chung-Yu. 2011. Intervention effects and wh-construals. *Journal of East Asian Linguistics* 21(1). 43–87. https://doi.org/10.1007/s10831-011-9080-5.

Yun, Jiwon. 2019. Meaning and prosody of wh-indeterminates in Korean. *Linguistic Inquiry* 50(3). 630–647. https://doi.org/10.1162/ling_a_00318.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin/New York: de Gruyter.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 4

https://doi.org/10.18148/zs/2026-2004

Martin Andreas Kvifte, Østfold University College/NTNU, Norway

The acquisition of verb and negation placement in L3 German

Abstract: The present study examines the acquisition of the verb-second (V2) word order and non-verb-adjacent negation with object scrambling in intermediate-level L3 German. 64 upper-secondary school students with L1 Norwegian and L2 English were recruited to perform three different elicitation tasks (acceptability judgments, element rearrangement, negation fill-in). The results show that the students generally have acquired the V2 word order, but struggle with non-verb-adjacent negation. Moreover, both previously learned languages seem to exert influence on the acquisition of German verb and negation placement. Whereas L2 English influences the students' verb placement, transfer from L1 Norwegian is present in their negation placement preferences. Finally, the students' verb placement accuracy differs depending on the linguistic context (adverbial complexity and verb type), suggesting that V2 is learned in a stepwise fashion.

Keywords: L3 acquisition, transfer, cross-linguistic influence, German verb and negation placement

Cite as: Kvifte, Martin Andreas. 2026. The acquisition of verb and negation placement in L3 German. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi.org/10.18148/zs/2026-2004
Received: 01 August 2025; **accepted:** 04 December 2025; **published:** 25 June 2026
Published by: publikon / dgfs (Linguistic Society of Germany)

1 Introduction

The verb-second (V2) word order has received much attention by both language teachers and researchers, and its successful acquisition is often considered a milestone in the process of learning Germanic languages other than English (Bohnacker & Westergaard 2024). When Scandinavian learners of German as a third language (L3) make word order mistakes, teachers often perceive this as negative influence from their second language (L2) English (Falk & Lindqvist 2022:163–166; Haukås 2016: 10), whereby violations of the V2 word order are referred to as "the English word order" (Beslagic 2024: 148–149). Yet, how and to what extent (morpho-) syntactic structures transfer from previously learned languages remains a controversial topic within the field of third language acquisition (L3A). Moreover, it has been suggested that rather than being one big rule, the V2 word order is learned stepwise, resulting in varying degrees of target-likeness depending on the linguistic context (Westergaard 2021a). The present study seeks to address these matters by examining the acquisition of L3 German finite verb and negation placement by L1 Norwegian learners with L2 English.

2 Verb and negation placement

Variation in verb and negation placement is here analysed as a result of different movement operations between the CP-TP-VP domains (Adger 2003; Wojtecka et al. 2011).[1] In Norwegian and German declaratives, finite verbs move from V (via T) to C (long movement), resulting in the V2 word order. Contrastingly, English (finite) lexical verbs cannot move out of the VP (Adger 2003; Pollock 1989), making them appear as the third constituent (V3) after topicalizations. Figure 1 below offers a simplified analysis of these movement differences.[2]

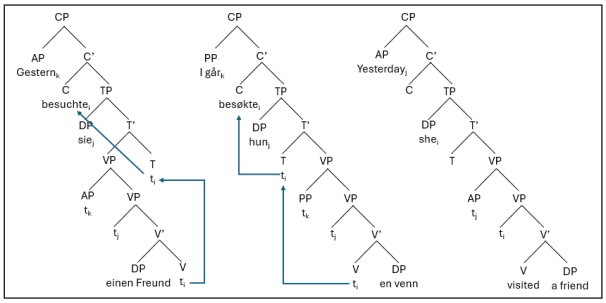

Figure 1: Verb movement in German, Norwegian and English

[1] Complementizer Phrase – Tense Phrase – Verb Phrase.

[2] To simplify the presentation in figure 1 and the following examples, adverbials and negation are analysed as adjoined to the VP in all three languages. The German VP and TP are head final due to German being an SOV language (Meisel 1997: 230).

Since English lexical verbs do not move, the same placement difference occurs in subject-initial declaratives with mid-sentential adverbials (1). English auxiliaries and copulas, however, may move to T (Adger 2003; Pollock 1989). This short movement becomes visible in main clauses with mid-sentential adverbials (2). Yet, after topicalizations, English auxiliaries appear in the third position since they do not move to C (3).

(1) a. [$_{CP}$ Ich **bestelle** [$_{TP}$ [$_{VP}$ oft einen Salat]]].
I order often a salad
b. [$_{CP}$ Jeg **bestiller** [$_{TP}$ [$_{VP}$ ofte en salat]]].
I order often a salad
c. [$_{CP}$ I] [$_{TP}$ [$_{VP}$ often **order** a salad]]].

(2) a. [$_{CP}$ Wir **werden** [$_{TP}$ [$_{VP}$ schließlich dieses Gespräch beenden]]].
We will eventually this conversation finish
b. [$_{CP}$ Vi **vil** [$_{TP}$ [$_{VP}$ omsider avslutte denne samtalen]]].
We will eventually finish this conversation
c. [$_{CP}$We [$_{TP}$ **will** [$_{VP}$ eventually finish this conversation]]].

(3) a. [$_{CP}$ Heute **werden** [$_{TP}$ wir [$_{VP}$ ein Buch lesen]]].
Today will we a book read
b. [$_{CP}$ I dag **skal** [$_{TP}$ vi [$_{VP}$ lese en bok]]].
Today will vi read a book
c. [$_{CP}$ Today, [$_{TP}$ we **will** [$_{VP}$ read a book]]].

Verb movement also influences the placement of sentential negation. Due to the verb's movement to C the negation appears after lexical verbs in German and Norwegian, whereas English must rely on *do*-support, leaving the lexical verb behind the negation as illustrated in (4).

(4) a. [$_{CP}$ Er **tanzt** [$_{TP}$ [$_{VP}$ nicht]]].
He dances not
b. [$_{CP}$ Han **danser** [$_{TP}$ [$_{VP}$ ikke]]].
He dances not
d. [$_{CP}$ He [$_{TP}$ does [$_{VP}$ not **dance**]]].

However, English auxiliaries and copulas move to T, resulting in the same verb-negation surface structure in all three languages.

(5) a. [$_{CP}$ Ich **bin** [$_{TP}$ [$_{VP}$ nicht dumm]]].
I am not stupid
b. [$_{CP}$ Jeg **er** [$_{TP}$ [$_{VP}$ ikke dum]]].
I am not stupid
c. [$_{CP}$I [$_{TP}$ **am** [$_{VP}$ not stupid]]].

In contrast to Norwegian and English, German sentential negation does not require finite-verb-adjacency.[3] In declaratives with objects, the unmarked negation placement is after the object, indicating that the object has been raised above the negation (*scrambling*) (Meisel 1997; Wojtecka et al. 2011). Thus, in declaratives with objects, negation appears in different positions relative to the lexical verb in all three languages (6): After the object in German, directly after the lexical verb in Norwegian, and in front of the lexical verb in English.[4]

(6) a. [$_{CP}$ Ich **mag** [$_{TP}$ [$_{VP}$ meine Stimme nicht]]].
I like my voice not
b. [$_{CP}$ Jeg **liker** [$_{TP}$ [$_{VP}$ ikke stemmen min]]].
I like not voice my
c. [$_{CP}$ I [$_{TP}$ do [$_{VP}$ not **like** my voice]]].

Yet, if the finite verb is an auxiliary, Norwegian and English have the same surface structure (7). In German, the negation appears behind the object, but in front of the lexical verb due to German being a Subject-Object-Verb (SOV/SXV) language.

(7) a. [$_{CP}$ Ich **habe** [$_{TP}$ [$_{VP}$ das Buch nicht gelesen]]].
I have the book not read
b. [$_{CP}$ Jeg **har** [$_{TP}$ [$_{VP}$ ikke lest boka]]].
I have not read book-the
c. [$_{CP}$ I [$_{TP}$ **have** [$_{VP}$ not read the book]]].

Finally, preverbal negation is the unmarked position in Norwegian imperative clauses (Garbacz & Johannessen 2014). This differs from English and German. In English, *do*-support makes the

[3] In this regard, one might raise the concern that the negation does not differ from other adverbials such as *mehrmals* ('many times'): *ich **habe** das Buch nicht gelesen* vs. *ich **habe** das Buch mehrmals gelesen*. In Norwegian and English, other adverbials may also appear separated from the finite verb (*jeg **har** lest boka flere ganger* / *I **have** read the book many times*), but this is not possible with sentential negation: *jeg **har** lest boka ikke / *I **have** read the book not. Hence, sentential negation behaves differently from other adverbials in the languages under study, and as depicted in Section 3 below, previous studies have found that non-verb-adjacent negation in particular poses challenges in the acquisition of German.

[4] Object pronouns may raise above the negation in Norwegian ('object shift'): *Jeg liker **den ikke***. The present study only examines full Determiner Phrases, which do not raise in Norwegian (Anderssen et al. 2018: 250).

negation occur post-verbally, yet in front of the lexical verb.[5] In German imperative clauses, the negation appears post-verbally with optional scrambling (8).

(8) a. [$_{CP}$ **Trink**] [$_{TP}$ [$_{VP}$ das Wasser <u>nicht</u>]]] / [$_{CP}$ **Trink** [$_{TP}$ [$_{VP}$ <u>nicht</u> das Wasser]]].
Drink the water not drink not the water.

b. [$_{CP}$ <u>Ikke</u> **drikk** [$_{TP}$ [$_{VP}$ vannet]]]. / [$_{CP}$ **Drikk** [$_{TP}$ [$_{VP}$ <u>ikke</u> vannet]]].[6]
Not drink water-the drink not water-the

c. [$_{CP}$ Do [$_{TP}$ not *pro* [$_{VP}$ **drink** the water]]].

3 Background

As mentioned in the introduction, L3 German teachers in Scandinavia link word-order errors to the students' L2 English. In line with this, some studies on L3A involving Germanic languages present evidence of transfer from L2 English (e.g. Bohnacker 2006; Bardel & Falk 2007). According to Falk & Bardel (2011: 77), the order of acquisition is of importance since second and third languages often share (cognitive) similarities "regarding age of onset, outcome, learning situation, metalinguistic knowledge, learning strategies and degree of awareness in the language learning process".[7] Falk & Bardel (2011) therefore suggest that the L2 overrides the L1 as a transfer source (the *L2 status factor*).[8]

However, the results of more recent studies question the validity of the L2 status. Instead, the authors argue for the availability of both previously learned languages (e.g. Dahl et al. 2022; Jensen et al. 2023; Stadt et al. 2020a). Stadt et al. (2020a) compared the acquisition of L3 German with L3 French after L1 Dutch and L2 English. The L3 German group made fewer V3 errors in sentences with mid-sentential adverbials than L3 French learners, indicating more transfer from L2 English in the L3 French group.[9] Stadt et al. (2020a) suggest that the typological closeness and structural similarity between Dutch and German caused an increased L1 influence in the L3 German group. Furthermore, Dahl et al. (2022) examined the acquisition of L3 German by

[5] Some learners might treat the contractions *don't* and *can't* as chunks, which could be interpreted as preverbal negation (Anderssen & Bentzen 2018: 4): "<u>*Don't drink*</u> *the water*". Moreover, it is assumed that *do*/*don't* move to C in imperatives since the *pro*-subject occupies the spec-TP position: [$_{CP}$ Don't [$_{TP}$ ~~you~~ *pro* [$_{VP}$ drink the water]]] (Han 2001: 303).

[6] Postverbal negation adjacent to the finite verb in Norwegian imperatives is dispreferred by younger generations (Garbacz & Johannessen 2014).

[7] According to Bardel & Sánchez (2017), these cognitive similarities disappear if the L2 proficiency level is very high and the language is largely learned through input without developing metalinguistic knowledge. Given the amount of extramural English input Norwegian students are exposed to, it has been argued that English no longer counts as a foreign language (Rindal 2019), meaning that English and German are acquired differently in the Norwegian school context.

[8] Some studies present evidence of the opposite: privileged L1 transfer (e.g. Hermas 2014).

[9] In contrast to English, French lexical verbs move to T, resulting in a surface V2 word order in subject-initial declaratives (Pollock 1989).

students with L1 Norwegian and L2 English. Judgment data from learners in their first, second, fourth and fifth year of German acquisition showed that beginners did not have any clear preference for either V2 or V3, rejecting the notion of a default L2 status. Increased years of instruction corresponded with more target-like verb placement. Since development towards target-like word order occurred earlier with mid-sentential adverbials, Dahl et al. (2022) argue that syntactic structures transfer stepwise ("property-by-property") as opposed to "wholesale".[10]

Largely compatible with these findings, the *linguistic proximity model* (LPM) (Westergaard et al. 2017) and the *scalpel model* (SM) (Slabakova 2017) predict that both previously acquired languages are available throughout L3 acquisition, whereby transfer takes place in small steps based on structural similarity detected through input.[11] Non-facilitative transfer may occur when the learner has not received sufficient input, or the input is misanalysed (e.g. due to complexity or ambiguity) and wrongfully perceived as similar to previously learned languages. Additionally, Busterud et al. (2023) suggest that non-facilitative transfer of verb placement is influenced by movement differences between the languages involved. If one of the previously acquired languages displays (lexical) verb movement, but the other not (e.g. L2 English), learners may be more conservative in performing movement operations in an L3 until sufficient evidence of verb movement has been established through input.

That languages are learned incrementally, is further supported by evidence indicating that transfer and the acquisition of verb placement vary depending on adverbial placement, adverbial complexity and verb type (Bohnacker & Westergaard 2024). Firstly, in addition to Dahl et al.'s (2022) finding of V2 being easier to learn in subject-initial declaratives L3 German, V2 seems more difficult to unlearn in subject-initial declaratives when acquiring English after a V2-language (Westergaard 2003). Secondly, studies on Norwegian as a heritage language show that ungrammatical V3 more often occurs after longer and less frequent adverbials, whereas V2 is more stable after short and frequent adverbials (Bohnacker & Westergaard 2024). Thirdly, V2 seems easier to learn, but more difficult to unlearn with auxiliaries than lexical verbs. Westergaard (2003) found that Norwegian learners of L2 English transferred more (ungrammatical) V2 in clauses with auxiliaries than with lexical verbs. Moreover, studies on the L2/L3 acquisition of negation placement in V2 languages show that postverbal negation occurs earlier with auxiliaries than lexical verbs (Bardel & Falk 2007; Meisel 1997). Thus, instead of being one big rule, Bohnacker & Westergaard (2024: 364) argue that the V2 word order consists of many small rules for different linguistic contexts.

[10] The *typological primacy model* predicts that, depending on overall (psycho-)typological similarity, either the L1 or the L2 grammar transfers completely ('wholesale' transfer) at the early stages of acquisition (Rothman 2015). Given the model's emphasis on lexical similarity as a trigger, a complete transfer of L1 Norwegian would be expected in German L3A with L1 Norwegian and L2 English, resulting in an instant acquisition of the V2 word order.

[11] This was earlier suggested in the *cumulative enhancement model*, which contrastingly predicts transfer to be only facilitative (Flynn et al. 2004).

Besides the reported differences between auxiliaries and lexical verbs, the above-mentioned finding of an accelerated V2 acquisition in subject-initial declaratives is of importance regarding German negation placement since postverbal negation requires V2. Accordingly, post-verbal negation should be learned relatively quickly, which is also what previous studies have found (Meisel 1997; Bardel & Falk 2007). However, the separation of the negation from the finite verb due to object scrambling seems more difficult to acquire for both L1 and L2 learners of German (Clahsen 1988; Wojtecka et al. 2011). Whereas children normally have acquired scrambling productively around the age of three, they have comprehension issues related to non-verb-adjacent negation for a prolonged period (Wojtecka et al. 2011: 232–239). According to Clahsen (1988: 23–24), non-verb-adjacent negation also emerges late in L2 German acquisition. Given the complexity of German negation placement, it is possible that learners avoid scrambling until they have received enough evidence of when to apply this movement operation. Otherwise, they would risk having to later unlearn it in certain contexts (Westergaard 2021a).

Finally, it has been hypothesized that non-facilitative transfer occurs more often in language production than in comprehension since the communication situation often forces learners to produce utterances beyond their proficiency level. They therefore resort to structures from previously learned languages (Westergaard 2021a). In a review of 71 transfer studies, Puig-Mayenco et al. (2020) found a correlation between production experiments and finding L2 transfer or influence from both previous languages. However, the results presented by Stadt et al. (2018, 2020a, 2020b) suggest that this changes with increased proficiency: Whereas L3 learners in their first year of instruction tended to make more errors in guided production (gap-filling) than in comprehension (grammaticality judgments), learners in their second or third years performed more target-like in the guided production task than in the judgment task.

4 Research design

As depicted in the above review, there is evidence that syntactic structures are acquired stepwise since they appear earlier in certain contexts than other. The present study seeks to examine the extent to which intermediate learners of L3 German with L1 Norwegian and L2 English have acquired German verb and negation placement in varying linguistic contexts. Moreover, there is uncertainty as to whether language production tests trigger more non-facilitative transfer than comprehension tests such as judgment tasks. Based on this, the following research questions are proposed:

1. To what extent have the students acquired German verb and negation placement?
2. To what extent does transfer affect the acquisition of verb and negation placement?
3. To what extent does adverbial placement and complexity, and verb type exert influence on the acquisition of German verb placement?
4. To what extent do (guided) production and judgment tasks yield different results?

It is predicted that:

a) the participants will perform more target-like on the V2 word order than negation with object scrambling. Whereas the former is shared by Norwegian and German, non-verb-adjacent negation is not available in Norwegian and English, and previous studies report difficulties related to the acquisition of non-verb-adjacent negation. Preverbal negation in declaratives is not expected since the participants are intermediate learners of German.
b) finite verb placement will be more target-like in subject-initial declaratives, in clauses with short and frequent adverbials, and with auxiliaries as opposed to finite lexical verbs.
c) in line with Stadt et al. (2020a), the participants will perform better in guided production than judgments.

It should be emphasized that the study does not aim to test the transfer models described above since this would require control groups (mirror-image or subtractive language groups) that are very difficult – if not impossible – to find (Westergaard 2021b: 508), but the results will nevertheless be discussed in light of the models. Whereas a mirror-image group would involve L1 English, L2 Norwegian, L3 German, subtractive language groups would require L1 Norwegian, L2 German and L1 English, L2 German. It is very difficult to recruit intermediate-level learners of German with such language backgrounds, who have received language instruction comparable to the participants in the present study.[12] Given the omnipresence of English in Norway, it is virtually impossible to recruit Norwegian students with L2 German.[13] Due to the difficulties involved in recruiting control groups, single group designs are common (cf. Puig-Mayenco et al. 2020: 49; Westergaard et al. 2023: 227–228). Despite its limitations, "this methodology can potentially point to important insights and directions for further research" (Westergaard et al. 2023: 227). Puig-Mayenco et al. (2020: 49) express that they "understand it is not always practical to find mirror-image groups. […] if this were a requirement, it would severely reduce the language pairing we would be able to study for obvious practical reasons."

4.1 Participants

To answer the research questions, 64 L1 Norwegian school students, age 16–18, were recruited to perform three tasks (cf. section 4.2) and fill out a language background questionnaire.[14] The

[12] Attempts were made to recruit school students abroad with L1 English, L2 German, but this proved unsuccessful due to the students either being complete beginners or having learned another foreign language prior to German (e.g. L1 English, L2 Spanish/French, L3 German).

[13] Bohnacker (2006) managed to recruit three L1 Swedish learners of L2 German, but these were old-age retirees. It might be possible to recruit similar participants in Norway, but such a group would have been small, and not directly comparable to the school students in the present study.

[14] 111 students performed the tasks, but 24 participants who did not complete all the tasks or had not filled out the language background questionnaire were excluded. Additionally, 23 participants were excluded for reporting

participants had just begun their fifth year of learning German and completed eleven years of compulsory English classes. An additional English acceptability judgment task was given to ensure that the students had acquired the structures under study in their L2. Furthermore, 23 L1 German school students, age 16–18, were recruited to validate the three German tasks.[15] The students were informed that participation was anonymous and voluntary. A written consent was collected from them. The study was reviewed and approved by the Norwegian Agency for Shared Services in Education and Research (Sikt).

4.2 Experiments

All data was collected digitally using Nettskjema (2024), which maintains a high level of privacy and security. The data collection was completed during school hours in the participants' classrooms. The tasks were untimed, but the students were instructed to answer the tasks spontaneously and to submit without checking their answers. The participants first performed the guided production (element rearrangement and negation fill-in), then the comprehension tasks (acceptability judgments, see Appendix). In the following, the design of the tasks is described in more detail.

4.2.1 Acceptability judgment task (AJT)

The two AJTs (one German and one English) consisted of 24 items: twelve grammatical sentences with twelve ungrammatical equivalents. The 16 items covering verb placement contained the same constituents (subject, verb, object, adverbial), but varied in terms of finite verb type (auxiliary, lexical verb), adverbial placement (topic, mid-sentence), and adverbial complexity (short/frequent, long/less frequent).[16] The eight sentences addressing negation could be divided into four categories (see Section 2): a) non-verb-adjacent negation with a lexical verb, b) non-verb-adjacent negation with an auxiliary, c) verb-adjacent negation with a lexical verb (object modified by an attributive subordinate clause), and d) verb-adjacent negation with an auxiliary.[17]

low grades in L3 German, for having additional home/heritage languages or learning an L4. In line with Dahl et al. (2022: 197–198), participants reporting rudimentary knowledge of additional languages (e.g. a few words, phrases, or sentences) were included in the study.

[15] 26 students performed the tasks, but three were excluded for having a different L1 than German.

[16] In topicalizations, subordinate clauses constituted the long/less frequent adverbials, whereas *gestern, morgen, heute* and *plötzlich* occured as short/frequent adverbials. In subject-initial sentences, the polysyllabic adverbials *möglicherweise, normalerweise, regelmäßig* and *hoffentlich* constituted the long/less frequent adverbials. Duden online (2025) classifies these adverbials into frequency class 3 out of 5, meaning that they have a frequency above 10 in a million in the Duden corpus. Contrastingly, the short, monosyllabic, mid-sentential adverbials belong to frequency classes 4 (*gern, oft*, above 100 in a million) or 5 (*dann*, above 1000 in a million). Duden online (2025) does not specify the frequency of the final mid-sentential adverbial, *jetzt*, but DWDS (2025) classifies it as very frequent.

[17] Since attributive subordinate clauses appear adjacent to their correlate (they may normally only be separated by an infinite verb part / verb particle), the unmarked negation placement in the superordinate clause is verb-adjacent

The participants were presented with one sentence at a time, and asked to decide whether they found the sentence to be "correct" or "wrong".[18] To ensure that the participants were judging the targeted structure, they were asked to correct sentences judged as "wrong" (Mackey & Gass 2015). This was generated automatically in Nettskjema at the end of the task. Since the students were performing three tasks and filling out a questionnaire, it was decided not to include filler sentences in the AJT. This may have increased the risk of the participants figuring out the target structures, but to extensive testing "can become tiresome and judgments may become unreliable" (Mackey & Gass 2015: 60).

4.2.2 Element rearrangement task (ERT)

The ERT involved sentence manipulation (Hawkins 2001: 23) and was designed to fit the story of a comic ("Vater und Sohn"), making the task semantically easier to understand. It consisted of eight sentences with constituents in random positions apart from the element to be fronted, which was marked in bold to elicit sentences with topicalizations in half of the sentences. In the other half, the subject was fronted, but violating the V2 word order was still possible due to adverbials (in one instance a negation) being part of the constituents. The task was not designed to systematically test different verb types (e.g. auxiliaries vs. lexical verbs). One of the topicalized sentences contained a long multi-word adverbial, to see if this triggered more V2 violations. The participants had to rearrange the constituents into their preferred order.

4.2.3 Negation fill-in task (NFT)

In the NFT, the participants were presented with eight sentences, which were to be negated. It resembled a gap-filling task, but without the gaps to examine the participants' preferred position without giving them any options that potentially could lead them in a specific direction. The sentences were designed using the same criteria as in the AJT with one exception: One of the sentences with potentially different negation in all languages was replaced by an imperative clause to test whether the participants to a larger degree would prefer preverbal negation in this sentence type (see Section 2).

4.3 Scoring and analysis

The data was initially scored in Microsoft Excel, where sentences with correct verb (V2) or unmarked negation placement (postverbal, object scrambling) were given one point. Non-target verb or negation placement received zero points. Other errors (morphology, orthography etc.) did

(Pittner & Berman 2010: 90–91): *Ich habe nicht das Gefühl, dass er mich wirklich liebt* ("I have not the feeling, that he me really loves").

[18] The dichotomous choice made the results easier to compare with the other tasks (binary scoring).

not impact the score. Scoring of the AJT followed the procedures of Falk & Bardel (2011), where the rejection of a non-target structure only received a point after its successful correction. If a participant rejected a sentence with non-target verb or negation placement, but made an incorrect correction (e.g. morphology, orthography) or no correction at all, zero points were given. If a grammatical sentence was corrected, but no changes were made to the verb or negation placement in the correction, it received a point despite having non-target verb conjugation, case markings or the like. In addition to the author, the entire dataset was scored by another researcher, resulting in an inter-rater reliability percentage of 99.46% (100% after comparison and discussion).

The scoring process described above provides insight into the participants' ability to produce/accept target-like structures, yet it does not necessarily account for whether the non-target productions or judgments can be explained as transfer from the participants' other languages or not. If an item addressing verb placement is given zero points, this does not automatically mean that V3 has been produced since non-target SXV-overgeneralizations, V4, missing constituents and insufficient corrections (e.g. "I don't know") also occur in the material. To examine to what extent non-target verb and negation placement coincided with potential transfer, the data was also analysed descriptively by coding the responses into different verb/negation-placement categories (e.g. V2, V3, SXV, preverbal/postverbal negation). After the initial scoring and coding, the data was further analysed with IBM SPSS Statistics.

5 Results

This section first depicts the total scores for both verb and negation placement combined. Then, descriptive analyses of verb placement and negation placement are presented individually. Figure 2 summarizes the main findings, which are presented in detail below.

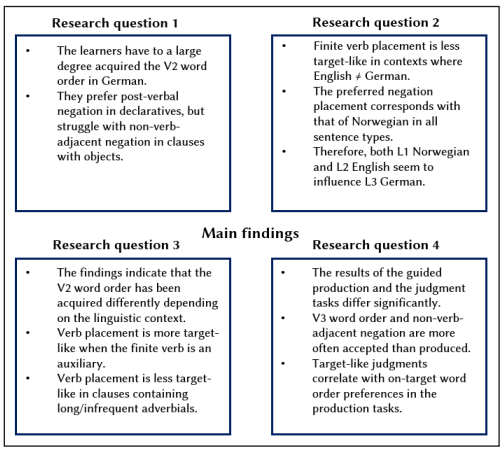

Figure 2: main findings

5.1 Total scores

The average total scores for the L3 German learners and the L1 German natives can be observed in Table 1. While the L3 German group scored 26.58 points on average, the standard deviation suggests there to be substantial differences in the individual scores of the participants. In the L1 German group, the standard deviation is much smaller and the average score at ceiling, indicating that the tasks provide a valid measurement of verb and negation placement proficiency. A t-test shows that the difference between the groups is significant ($p < .001$, d = 2.811).[19]

Table 1: Total scores – L3 group vs. L1 group

Group	N	Mean	Standard deviation
L3 German	64	26.58/40 (66.5%)	4.902
L1 German	23	38.57/40 (96.5%)	1.199

Furthermore, the mean difference between the L3 German and the L2 English AJTs (Table 2) is significant ($p < .001$, d = 2.416). The scores in the English AJT are near ceiling, suggesting that the students to a large degree have acquired English verb and negation placement, making these properties available for potential transfer.[20]

Table 2: Total AJT scores L3 vs. L2

Group	N	Mean	Standard deviation
L3 German	64	15.05/24 (62.7%)	3.583
L2 English	64	22.03/24 (91.8%)	1.968

Both means and standard deviations indicate that the participants are more insecure in the German AJT compared to the English one. Table 3 illustrates that the participants are generally more insecure when judging ungrammatical sentences, but they rarely reject correct verb placement. When judging negation placement, they struggle both with grammatical/unmarked and ungrammatical/marked sentences, suggesting that negation is more difficult for them. The score differences are significant (χ^2 (3) = 448.968, p <.001) with a large effect size (V = .541).[21]

[19] Interpretation of effect sizes (Cohen's d): 0.2 = small, 0.5 = medium, 0.8 = large (Geher & Hall 2014: 206).

[20] Some participants occasionally reject grammatical sentences with topicalizations (30/256) (e.g. *Later, she studied classical music*) and replace them with subject-initial sentences (*She studied classical music later*). Since these participants always accept grammatical V3 in clauses with mid-sentential adverbials and mostly reject ungrammatical V2, these instances are understood as a dispreference for topicalizations rather than a rejection of the finite verb placement. If these instances were counted as a rejection of the verb placement, the mean score would drop to 21.6/24 (90%).

[21] See Kim (2017) for the interpretation of Cramer's V effect sizes relative to the degree of freedom.

Table 3: On-target judgment frequencies relative to item type and structure in the L3 German AJT

	Verb placement		Negation placement	
Item type	Grammatical	Ungrammatical	Grammatical	Ungrammatical
On-target frequency	95.7% (490/512)	43.8% (224/512)	68.8% (176/256)	28.5% (73/256)

*All groups differ significantly from each other.

5.2 Verb placement

As mentioned in Section 4.3, the total scores do not account for the extent to which non-target productions or judgments can be related to previously learned languages. The following paragraphs will therefore contain a descriptive overview of the participants' preferences.

Table 4 shows that the participants have to a large degree acquired V2 in German, yet 26.9% of the results are non-target. Whereas ungrammatical V3 constitutes most of the students' off-target judgments/productions, they also produce non-target SXV to a small extent – a possible overgeneralization of the word order in German subordinate clauses.

Table 4: Finite verb placement in the ERT and the AJT

Finite verb placement	Frequency	Examples
V2 (target word order)	73.1% (1123/1536)	a) Ich bestelle oft einen griechischen Salat. *I order often a Greek salad.*
V3	20.2% (310/1536)	b) *ich oft bestelle einen griechischen Salat. *I often order a Greek salad.*
SXV	3.4% (52/1536)	c) *Ich oft einen Salat griechischen bestelle. *I often a Greek salad order.*
Other[22]	3.3% (51/1536)	d) * Ich oft einen Salat bestelle griechischen. *I often a salad order Greek.*

5.2.1 Verb placement production (ERT) vs. comprehension (AJT)

To compare non-facilitative transfer in production and comprehension, Figure 3 depicts the ungrammatical V3 scores from the ERT and the AJT individually.[23]

[22] The category "other" refers to a) missing AJT-corrections or the answer "I don't know" (29/51); b) odd verb placement such as V4 (14/51); c) inseparability between V3 and SXV-overgeneralizations (5/51): *Der Sohn dort liegt und liest (*The son there lies and reads*); d) inseparability between V2 and V3 (2/51): ?Darum der Sohn soll der Sohn den Vater suchen (*Therefore, the son shall the son the father seek/look for*); and e) missing verb (1/51).

[23] One ERT sentence ("Der Sohn liegt dort und liest") is not included in the analysis because it cannot be determined whether non-target productions have V3 or SXV word order: "*Der Sohn dort liegt und liest".

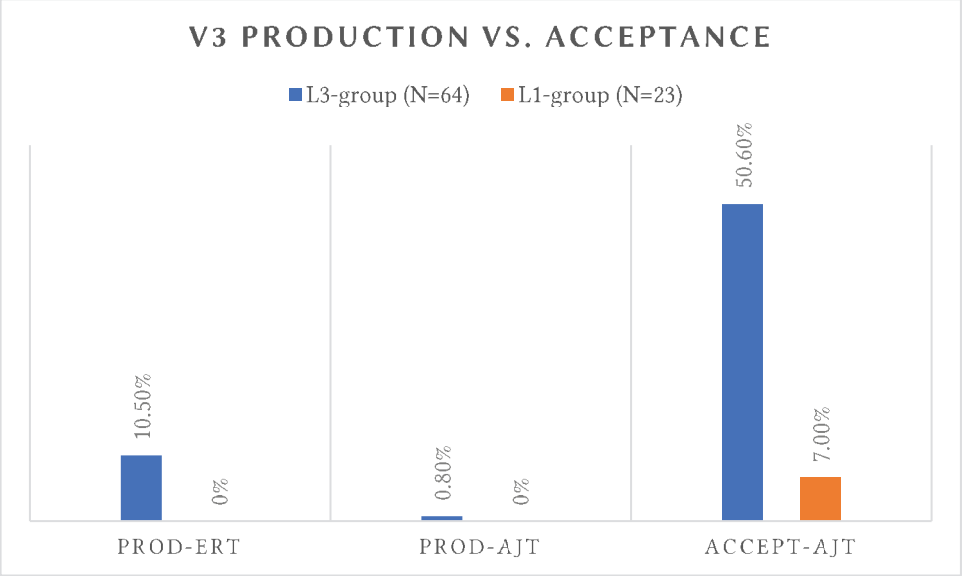

Figure 3: V3 production vs. acceptance

Since the AJT involves correction, the task has an element of production to it. The AJT scores have therefore been divided into two categories: acceptance of V3 (ACCEPT-AJT) and production of V3 (PROD-AJT), the latter meaning that the participant has rejected on-target V2 and produced V3 in the correction attempt. This rarely occurs: Out of the 512 judgments, V2 was rejected and changed to V3 only four times. In comparison, the acceptance of V3 sentences in the AJT is substantially higher (50.6%, 259/512). This also stands in contrast to the average V3 production in the ERT (10.5%, 47/448). The differences are significant (χ^2 (2) = 425.238, p < .001, V = .537).[24] Moreover, there is a weak positive correlation (r = .186, p < .001) between V2 production in the ERT and the ability to reject and correct ungrammatical V3 in the AJT.[25] To summarize, the V3 word order is more likely to be accepted than to be produced. This pattern also occurs in the L1 group as they produce no V3 sentences but, to a small extent, accept them (7%, 13/184), indicating yes-bias in the AJT.[26]

5.2.2 Verb placement in the acceptability judgment task

The AJT items testing verb placement included either topicalized or mid-sentential adverbials. A comparison of these two sentence types (Table 5) did not reveal any differences.[27]

[24] Post-hoc test: All groups differ significantly from each other.
[25] Interpretation of Pearson's r: 0.0–0.3 = weak, 0.3–0.7 = moderate, 0.7–1.0 = strong (Geher & Hall 2014: 70).
[26] In the L3-group, the average acceptance rate in the AJT is at 15.3/24 (with 12 being a perfect discrimination between grammatical and ungrammatical sentences), which suggests a small yes-bias.
[27] Moreover, leaving out heavy adverbials and auxiliaries (potential confounding variables), did not yield any significant differences.

Table 5: Verb placement in the L3 German AJT

	Adverbial placement		**Adverbial type**		**Finite verb type**	
	Topic	Mid-sentence	Long adverbial	Short adverbial	Lexical	Auxiliary
V2	69.5%[a] (356/512)	69.9%[a] (358/512)	65%[a] (333/512)	74.4%[b] (381/512)	66.5%[a] (340/512)	73%[b] (374/512)
V3	26%[a] (133/512)	25.4%[a] (130/512)	30.3%[a] (155/512)	21.1%[b] (108/512)	27.5%[a] (141/512)	23.8%[a] (122/512)
SXV	1.8%[a] (9/512)	1.8%[a] (9/512)	2.55%[a] (13/512)	1%[a] (5/512)	3.3%[a] (17/512)	0.2%[b] (1/512)
Other	2.7%[a] (14/512)	1.9%[a] (15/512)	2.15%[a] (11/512)	3.5%[a] (18/512)	2.7%[a] (14/512)	3%[a] (15/512)

*Different letters (e.g. [a] - [b]) denote significant differences.

The complexity of the adverbials, however, influences verb placement scores as heavy adverbials (topicalized subordinate clauses, (9), polysyllabic mid-sentential adverbials, (10) correspond with a higher average V3 acceptance than short/frequent adverbials (topicalized single words, (11–12), monosyllabic mid-sentential adverbials, (13–14). The difference is significant (χ^2 (3) = 16.871, p < .001, V = .128).

(9) *Weil sie Journalistin ist, sie schreibt viele Artikel.
Since she journalist is, she writes many articles.

(10) *Ich normalerweise trinke ein großes Bier.
I normally drink a large beer.

(11) *Heute er präsentiert sein neues Buch.
Today he presents his new book.

(12) *Gestern er hat seinen Großvater besucht.
Yesterday he has his grandfather visited.

(13) *Ich oft bestelle einen griechischen Salat.
I often order a Greek salad.

(14) *Du dann musst ein neues Leben beginnen.
You then must a new life begin.

Another difference was found between lexical verbs (11), (13) and auxiliaries (12), (14): χ^2 (3) = 17.248, p < .001, V = .130. The students perform more target-like (more V2) when the finite verb is an auxiliary. With lexical verbs, ungrammatical V3 is more often accepted, and when rejected, more often replaced with ungrammatical SXV-overgeneralisations compared to auxiliaries. Moreover, the long/less frequent adverbials (9,) (10) emerge as a potential confounding variable: When comparing ungrammatical sentences with long/less frequent adverbials, the acceptance of V3 is equally high regardless of verb type (Table 6).

Table 6: V3 acceptance relative to adverbial type in the AJT

	Long/infrequent adverbials		**Short/frequent adverbials**	
Verb type	Lexical verb	Auxiliary verb	Lexical verb	Auxiliary verb
V3	60.2%[a] (77/128)	60.2%[a] (77/128)	47.7%[a] (61/128)	34.4%[b] (44/128)

*Different letters (e.g. [a] - [b]) denote significant differences.

Contrastingly, if only the ungrammatical sentences with short/frequent adverbials are considered, the students more often accept ungrammatical V3 with lexical verbs than with auxiliaries (χ^2 (1) = 4.666, p = .031, V = .135). In the L2 English AJT, a similar pattern can be observed. The participants are better at rejecting and correcting ungrammatical sentences with lexical verbs.[28] In other words, the participants average V3 scores are higher in sentences with lexical verbs in both the English and the German AJT.[29] Hence, lexical verbs correspond with more target-like performance in L2 English, but more errors in L3 German.

5.2.3 Verb placement in the element rearrangement task

In contrast to the AJT, verb placement differences in relation to adverbial placement were found in the ERT: χ^2 (3) = 8.231, p = .041, V = .146.[30] The post-hoc tests show that the students produce more ungrammatical V3 in declaratives with topicalized adverbials (Table 7), but this does not mean that their productions are more target-like in clauses with mid-sentential adverbials. Instead, they make other mistakes such as SXV-overgeneralizations or placing the infinite verb in the second position (**Darum suchen der Sohn soll den Vater / 'therefore, seek the son shall the father'*).

[28] χ^2 (1) = 12.034, p < .001, V = .153.

[29] In clauses with mid-sentential adverbials and auxiliaries, V2 word order is grammatical in English. When analysing clauses with topicalized adverbials only, the (target-like) V3 scores are significantly higher with lexical verbs than auxiliaries (92.2% vs. 75.8%): χ^2 (1) = 12,807 p < .001, V = .224.

[30] ERT sentences h) and c) (see Appendix) were omitted from the analysis since h) contains a heavy adverbial and c) does not distinguish between V3 and SXV.

Table 7: Verb placement preferences in the ERT

	Adverbial placement		Adverbial type		Finite verb type	
	Topic	Mid-sentence	Heavy adverbial	Light adverbial	Lexical	Auxiliary
V2	78.6%[a] (151/192)	80.2%[a] (154/192)	70.3%[a] (45/64)	76.6%[a] (98/128)	78.9%[a] (202/256)	80.5%[a] (103/128)
V3	11.5%[a] (22/192)	4.7%[b] (9/192)	25%[a] (16/64)	12.5%[b] (16/128)	7.8%[a] (20/256)	8.6%[a] (11/128)
SXV	7.3%[a] (14/192)	9.4%[a] (18/192)	3.1%[a] (2/64)	10.9%[a] (14/128)	12.1%[a] (31/256)	0.8%[b] (1/128)
Other	2.6%[a] (5/192)	5.7%[a] (11/192)	1.6%[a] (1/64)	0%[a] (0/128)	1.2%[a] (3/256)	10.2%[b] (13/128)

*Different letters (e.g. [a] - [b]) denote significant differences.

Similar to the AJT, there are differences associated with the adverbial type as the ERT sentence with a long adverbial involves more V3 (χ^2 (3) = 9.349, p = .025, V = .221).[31] Still, the sentences with short adverbials are not more grammatical due to a higher SXV-overgeneralization frequency. Finally, differences regarding verb type were found (χ^2 (3) = 29.763, p< .001, V = .278).[32] The post-hoc tests revealed that these differences are not associated with V2 or V3: Whereas the students produce more SXV-overgeneralizations in declaratives with lexical verbs, they more often mix up finite and infinite verb forms in sentences with auxiliaries. To summarize, no differences in target-like V2 production related to adverbial placement, adverbial type or verb type were found, but these conditions yielded significantly different ungrammatical outcomes.

5.3 Negation placement

As described in Section 2, verb and negation placement are related to one another: Postverbal negation requires verb movement. Table 8 shows that the students predominantly choose postverbal negation, suggesting that such movement on average is in place.

5.3.1 Unmarked negation placement in declaratives

German differs from Norwegian and English in that unmarked German negation is not adjacent to the finite verb in sentences with objects (examples c and d in Table 9). Most participants have seemingly not acquired this feature as negation placement accuracy differs relative to whether

[31] As the ERT contains just one sentence with a heavy adverbial (sentence h), only sentences with topicalizations and lexical verbs were compared to keep things equal.

[32] ERT sentences h) and c) were not included for the same reasons mentioned above.

the unmarked negation placement is verb-adjacent or not (χ^2 (3) = 343.100, p < .001, V = .598). Post-hoc tests exhibit that their negation placement accuracy is higher when the unmarked position is directly after the finite verb (examples a and b in Table 9).[34]

Table 8: Negation placement frequencies (regardless of whether on- or off-target)

Negation placement	Frequency	Examples
Postverbal (adjacent to the finite verb)	77.6% (795/1024)	a) Wir können <u>nicht</u> sicher sein. *We can not be sure.*
Postverbal + raised obj. (not adjacent to the finite verb)	11.7% (120/1024)	b) Er wäscht seine Hände <u>nicht</u>. *He washes his hands not.*
Preverbal (adjacent to the finite verb)	3.6% (37/1024)	c) *Er nicht wäscht seine Hände. *He not washes his hands.*
Other[33]	7.1% (72/1024)	d) *Wir können sicher <u>nicht</u> sein. *We can sure not be.*

Table 9: Negation placement accuracy in the NFT and the AJT (on-target scores)

Unmarked negation placement	Accuracy	Examples
Verb-adjacent (lexical finite verb)	82.8% (212/256)	a) Er tanzt nicht. *He dances not.*
Verb-adjacent (auxiliary)	82% (210/256)	b) Sie kann <u>nicht</u> schwimmen. *She can not swim.*
Non-verb-adjacent (lexical finite verb)	17.2% (33/192)[35]	c) Ich verstehe die Aufgabe <u>nicht</u>. *I understand the task not.*
Non-verb-adjacent (auxiliary)	27.7% (71/256)	d) Ich habe das Buch <u>nicht</u> gelesen. *I have the book not read.*

[33] This category refers to a) ungrammatical scrambling of predicative adjectives (35/71, cf. example d) in table 8); b) missing AJT-corrections/"I don't know" (20/72); c) inseparability between verb-adjacent preverbal negation and SXV-overgeneralization (4/72): *er nicht tanzt (*he not dances*); d) overgeneralization of sentence-final negation (4/72): *Ich habe das Buch gelesen nicht (*I have the book read not*); e) SXV-overgeneralization (3/72): *Ich nicht dumm bin (*I not stupid am*); f) missing negation (3/72); g) negation in odd positions (2/72): *Ich habe das nicht Gefühl, dass er mich wirklich liebt (*I have the not feeling, that he really loves me*) and h) transforming the sentence into an infinitive phrase (1/72): Nicht das Wasser trinken (*Not the water drink*).

[34] In the L1 group, 8/23 participants unexpectedly produce non-verb-adjacent negation in an NFT-sentence with a relative clause, splitting this clause from its correlate: "Ich habe die Antwort <u>nicht</u>, die du willst" (*I have the answer not, that you want*). It has been suggested that the negation could belong to the constituents that may appear between attributive clauses and the correlate (Helbig & Buscha 2000: 232), but this is considered controversial (Pittner & Berman 2010: 91). Nevertheless, this shows that the L1 group have a stronger preference for object scrambling than the L3 learners, who all produce verb-adjacent negation in the same sentence.

[35] The imperative sentence is not included as it has optional scrambling.

5.3.2 Negation with scrambling in production (NFT) vs. comprehension (AJT)

There are differences between the NFT and the AJT with regards to non-verb-adjacent negation. Figure 4 shows that the L3 learners tend not to produce such negation in the NFT (PROD-NFT). To reject marked verb-adjacent negation in favour of unmarked non-verb-adjacent negation occurs even less often (PROD-AJT), but the L3 group is more willing to accept negation with object scrambling.[36] The L1 group, in comparison, consistently produce/accept non-verb-adjacent negation, but they do not always reject marked verb-adjacent negation (PROD-AJT). Again, these findings indicate a yes-bias in the AJT. However, the L3 group to a large degree also reject unmarked non-verb-adjacent negation and replace it with verb-adjacent negation 47.6% (61/128). Moreover, preferring non-verb-adjacent negation in the AJT correlates positively with the production of such negation in the ERT (r = .601, p < .001).

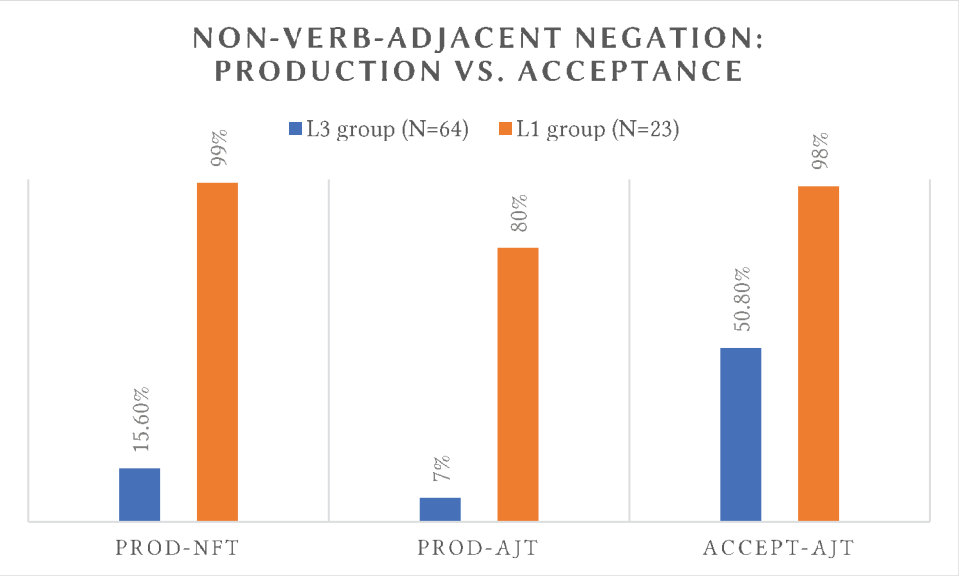

Figure 4: Frequencies of negation placement with object scrambling.[37]

5.3.3 Preverbal negation and imperatives

The students seldom produce preverbal negation. Besides 23 students accepting a sentence with preverbal negation in the AJT, only once is postverbal negation rejected in favour of preverbal negation. Among the declaratives in the NFT, preverbal negation is suggested twice. This is different in the imperative sentence in the NFT.

[36] χ^2 (2) = 79.580, p < .001, V = .421. Post-hoc test: Only the difference between ACCEPT-AJT and the other categories is significant.

[37] The imperative is not included as object-raising is optional in this clause type.

Table 10: Negation placement in the NFT

	L3 Group		**L1 Group**	
	Declarative	Imperative	Declarative	Imperative
Postverbal (Non-verb-adjacent)	6.2%[a] (4/64)	25%[b] (16/64)	100%[c] (23/23)	52.2%[b] (12/23)
Post verbal (verb-adjacent)	93.8%[a] (60/64)	56.3%[b] (36/64)	0%[c] (0/23)	47.8%[b] (11/23)
Preverbal (verb-adjacent)	0%[a] (0/64)	17.2%[b] (11/64)	0%[a] (0/23)	0%[a] (/23)
Other[38]	0%[a] (0/64)	1.5%[a] (1/64)	0%[a] (0/23)	0%[a] (0/23)

*Different letters denote significant differences.

Table 10 compares the L3 and L1 groups' placement preferences in two NFT-sentences, which are similar in that they both feature a single lexical verb and an object. They differ in that one of them is an imperative (17), the other a declarative (18). The comparison revealed negation placement and group differences (χ^2 (9) = 97, p < .001, V = .432).

(15) a. Trink das Wasser nicht.
Drink the water not.
b. Trink nicht das Wasser
Drink not the water.
c. *Nicht trink das Wasser.
Not drink the water.

(16) a. Ich mag meine Stimme nicht. *(unmarked)*
I like my voice not.
b. Ich mag nicht meine Stimme. *(marked)*
I like not my voice.
c. *Ich nicht mag meine Stimme.
I not like my voice.

Whereas the L3 learners produce no preverbal negation in the declarative (18c), eleven students suggest preverbal negation in the imperative (17c) – a negation placement that happens to be the unmarked position in Norwegian imperatives.

[38] The utterance was converted into an infinitive phrase ("*Nicht das Wasser trinken*").

6 Discussion

The present study sought to examine the following research questions:

1. To what extent have the students acquired German verb and negation placement?
2. To what extent does transfer affect the acquisition of verb and negation placement?
3. To what extent does adverbial placement and complexity, and verb type exert influence on the acquisition of German verb placement?
4. To what extent do (guided) production and judgment tasks yield different results?

The results show that the participants on average have acquired the V2 word order, but also produce and (to a larger degree) accept non-target V3 word order. The extent to which the participants allow or produce V3 varies under different syntactic conditions (further details below). Regarding negation, the learners have a strong preference for verb-adjacent post-verbal negation, indicating that the German non-verb-adjacent negation in declaratives with objects is difficult to acquire. There are many possible reasons for this. Firstly, the German input is complex as learners encounter both verb-adjacent and non-verb-adjacent postverbal negation, which may cause misanalyses of the input, leading to a transfer of verb-adjacent negation. Secondly, neither Norwegian nor English have scrambling in negated declaratives. To avoid potential unlearning, the learners may hesitate to perform this movement until they, through input, have established when to scramble in German (Westergaard 2021a).

Preverbal negation is virtually non-existent in the students' productions, which was expected given that they were intermediate learners of German. Such negation placement predominantly occurred in an imperative clause. This is interpreted as transfer from L1 Norwegian as preverbal negation is the unmarked position in Norwegian declaratives (Garbacz & Johannessen 2014). One could argue that the preverbal negation in **Nicht trink das Wasser* was transferred from English *Don't drink the water* because the participants may have treated *don't* as a chunk equal to *nicht*. However, this is unlikely since they did not produce preverbal negation in other contexts where the negation appears in front of the lexical verb in English (e.g. **Ich nicht mag meine Stimme* vs. *I don't like my voice*). Instead, the participants' negation placement preferences in German above all match those of Norwegian in all sentence types, suggesting that their L1 Norwegian exerts influence on their L3 German – an influence that is facilitative in contexts where German has verb-adjacent negation, and non-facilitative when German has object scrambling.

While the acceptance of ungrammatical V3 indicates transfer from L2 English, L1 Norwegian seems to exert a strong influence on L3 German negation placement. This suggests that both previously learned languages are available for transfer and – as demonstrated by

the L1 Norwegian influence on negation placement – it can be both facilitative and non-facilitative. As such, the results are compatible with the LPM (Westergaard et al. 2017) and the SM (Slabakova 2017). In this regard, the lack of the control groups described in section 4 constitutes a limitation of the present study. Without control groups it is difficult to determine whether other factors than transfer, such as general learning effects, may have influenced the results. The above-mentioned models may therefore not be directly tested. Still, the results clearly indicate that transfer plays a role. For instance, the tests included both sentences where English has the same verb placement as German and Norwegian (auxiliaries), and sentences where English differs from German and Norwegian (lexical verbs). Since the students were more target-like at judging sentences, where all three languages have the same verb placement (auxiliaries), but struggled more with sentences in which English differs from German and Norwegian (lexical verbs), this is an indication of influence from English. Moreover, the participants' preferred negation placement corresponds with that of Norwegian in the different sentence types.

The compatibility of the findings with the LPM and the SM reaches beyond that of influence from both previously learned languages. These models hypothesize that the L3 grammar is built incrementally in a stepwise fashion. The findings of the present study indicate that the V2 word order has been acquired to a different extent depending on the adverbial complexity and verb type, supporting "analyses that treat V2 word order as the result of many smaller rules" (Bohnacker & Westergaard 2024: 364). Firstly, the participants' verb-placement preferences are more target-like in clauses with short/frequent adverbials. Similar to the results in studies on Norwegian as a heritage language (Bohnacker & Westergaard 2024), more V3 occurred after topicalizations with long/less frequent adverbials. A new finding in the present study is that this also occurs in subject-initial clauses with long/less frequent mid-sentential adverbials.

Secondly, the participants make more accurate judgments in the German AJT when the finite verb is an auxiliary as opposed to a lexical verb. Simultaneously, they more often accept non-target auxiliary V2 in the English AJT. Thus, a similar pattern emerges in both tests: V2 is more likely to be accepted if the finite verb is an auxiliary. Westergaard (2003), who found a similar pattern in the L2 English of young Norwegians, suggests that learners may be misled by the fact that English auxiliaries behave differently (short movement in declaratives, long movement in *wh*-questions), causing the misinterpretation that English auxiliaries have long movement also after topicalizations in declaratives (Westergaard 2003: 95). In the present study, this pattern may have been transferred from L2 English into the participants L3 German, which would explain why they perform equally well on clauses with auxiliaries regardless of whether they are subject-initial or have topicalized adverbials. The larger acceptance for V3 with lexical verbs supports Busterud et al.'s (2023) suggestion that L3 learners are more

conservative in performing movement operations if a previously learned language (English) does not allow such movement.

However, no clear differences related to adverbial placement were found. This diverges from the results of Dahl et al. (2022) but may be explained by the participants' different proficiency levels as Dahl et al. (2022) included both beginners and intermediate learners in their analysis, whereas this study only tested intermediate learners.

Finally, guided production and acceptability judgments yielded different results. In line with the results of Stadt et al. (2020a), there is less ungrammatical V3 in the production task (ERT) compared to the comprehension task (AJT). This seemingly goes against the notion that production triggers more non-facilitative transfer (Westergaard 2021a), but regarding negation placement, the L1 influence is stronger in the production task (NFT) than in the AJT. In both cases, the results do not correspond with Puig-Mayenco et al.'s (2020) finding of more L2 transfer in productions tasks. The higher acceptance rates in the AJT are most likely influenced by a certain yes-bias. Also, Stadt et al. (2020b: 253) relate the higher acceptance rates in their judgment tasks to possible test effects, suggesting that judgments are demanding since they require both reading skills, the ability to focus and morphosyntactic knowledge.[39] Their participants reported finding guided production easier. The tendency to accept both grammatical and ungrammatical sentences could be a sign of insecurity related to a generally low target-language proficiency: Whereas the average acceptance rate in the L3 German AJT (in the present study) is at 15.3/24 (with 12 being a perfect discrimination between grammatical and ungrammatical items), the acceptance rate in the L2 English AJT is 12.75/24, where they scored almost at ceiling. The availability of two previously learned languages with different verb placement patterns may add to the insecurity – maybe above all when learners are forced to decide whether a sentence is grammatical or not.

In the AJT, the participants reject both grammatical and ungrammatical negated clauses. Especially grammatical non-verb-adjacent negation is rejected and replaced by verb-adjacent negation. Contrastingly, they virtually always accept grammatical verb placement. Hence, they struggle more with negation than verb placement, which may be due to the complexity of negation placement and object scrambling not being available in either previously learned languages. The ability to produce scrambling in the NFT correlates positively with non-verb-adjacent negation accuracy in the AJT, and there is a similar correlation between V2 production in the ERT and ability to reject and correct ungrammatical verb placement in the AJT. The findings in the AJT therefore cannot only be described as the result of a test effect.

[39] Stadt et al. (2020b: 254) note that they did not randomize the items, which may have influenced the results. This is also the case for the present study. Moreover, an anonymous reviewer suggested randomizing the order of the tasks to ensure that not all participants would perform the production task before the comprehension task and vice versa. Future studies should take this into consideration.

7 Conclusion

The results of the present study suggest that the V2 word order has been largely acquired by intermediate L3 learners of German with L1 Norwegian and L2 English. Transfer of the V3 word order from English seems present, but to a low degree (in production), whereby target-likeness differs depending on adverbial complexity and verb type, indicating that the acquisition of V2 is a stepwise, context-dependent process. Furthermore, post-verbal negation in declaratives is in place, but the learners struggle with German non-verb-adjacent negation. Above all, they prefer verb-adjacent negation placement similar to Norwegian. Both previously learned languages seem therefore to exert influence on L3 German. Finally, guided production and judgement tasks yielded different results, which is likely a consequence of the participants being insecure when judging sentences, leading to a yes-bias. The results could perhaps have been clearer if the grammatical and ungrammatical sentences were presented pairwise, making the participants choose between them. Alternatively, future studies could employ a different comprehension task such as self-paced reading.

Appendix

Element rearrangement task (ERT)

- Bringen Sie die Wörter in die richtige Reihenfolge. Beginnen Sie mit dem **fett gedruckten Glied**.

a) [essen] **[Der Vater]** [nicht] [darf] [ohne den Sohn]
b) [der Vater] [ins Kinderzimmer] **[Deshalb]** [geht]
c) [liegt] **[Der Sohn]** [und] [dort] [liest]
d) **[Der Vater]** [den Sohn] [zuerst] [bittet] [, zum Essen zu kommen]
e) [schaut] [auf das Buch] **[Dann]** [der Vater]
f) [auf den Vater] [jetzt] **[Die Mutter und der Sohn]** [warten]
g) [suchen] [den Vater] [der Sohn] **[Darum]** [soll]
h) [der Sohn] **[Zurück in seinem Kinderzimmer]** [den Vater] [findet] [lesend auf dem Fußboden]

Quelle: https://de.wikipedia.org/wiki/Datei:Vater_und_Sohn_-_Der_Schmoeker.png

Negation fill-in task (NFT)

- Setzen Sie das Wort „**nicht**“ in den Satz hinein.

a) Ich mag seine Stimme.
b) Sie kann schwimmen.
c) Er tanzt.
d) Ich habe die Antwort, die du willst.*
e) Wir können dieses Thema diskutieren.
f) Sie hat den Film gesehen.
g) Trink das Wasser!
h) Ich bin dumm.

*Verneinen Sie den unterstrichenen Satzteil mit „nicht“

Acceptability judgment task (AJT)

- Finden Sie den Satz grammatisch richtig oder falsch?

a) Ich baue jetzt meine eigene Garage.
b) Ich verstehe nicht die Aufgabe.
c) Gestern er hat seinen Großvater besucht.
d) Weil ich Allergiker bin, esse ich keine Nüsse
e) Wir können nicht sicher sein.
f) Du dann musst ein neues Leben beginnen.
g) Morgen wasche ich meinen roten Audi.
h) Ich habe das Buch nicht gelesen.
i) Wenn ich groß bin, ich möchte meine Freundin heiraten.
j) Wir können gern eine lange Pause machen.
k) Heute er präsentiert sein neues Buch.
l) Er kann Pilot nicht werden.
m) Weil sie Journalistin ist, sie schreibt viele Artikel.
n) Du hast möglicherweise die falsche Methode benutzt.
o) Sie hat nicht die Rechnung bezahlt.
p) Ich normalerweise trinke ein großes Bier.
q) Er wäscht seine Hände nicht.
r) Wir finden hoffentlich ein gutes Hotel.
s) Ich habe nicht das Gefühl, dass er mich wirklich liebt.
t) Sie regelmäßig hat ein dummes Computerspiel gespielt.
u) Plötzlich haben wir eine Stimme gehört.
v) Wir nicht haben den Eindruck, dass diese Lösung funktioniert.
w) Wenn du fertig bist, kannst du deinen Vater wecken.
x) Ich oft bestelle einen griechischen Salat.

L2 English test

Acceptability judgement task

- Do you think the following sentences are grammatically correct?

a) Later, she studied classical music.
b) He likes not his own father.
c) Because he is older, he should get the bigger bedroom.
d) Today, will we read an interesting book.
e) She does not understand the problem.
f) When he is angry, he says horrible things.
g) She still visits her old grandmother.
h) She lives not in Paris anymore.
i) She occasionally teaches academic writing.
j) I would then repeat the question.
k) He can find his mother not.
l) Yesterday, played he computer games.
m) We eventually will finish this conversation.
n) Later, you can watch the beautiful sunset.
o) She has not done her homework.
p) Because she is French, will she study literature.
q) He repaired then the washing machine.
r) That is not the man I saw.
s) We soon will discuss this issue.
t) I do not want to talk about it.
u) He drinks regularly fresh orange juice.
v) You not are the right person for this job.
w) When I am hungry, eat I a large sandwich.
x) She can finally write her own name.

References

Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford University Press.

Anderssen, Merete & Kristine Bentzen. 2018. Different outcomes in the acquisition of residual V2 and do-support in three Norwegian-English bilinguals: Crosslinguistic influence, dominance and structural ambiguity. *Frontiers in Psychology* 9. 21–30. https://doi.org/10.3389/fpsyg.2018.02130.

Anderssen, Merete, Kristine Bentzen, Guro Busterud, Anne Dahl, Bjørn Lundquist & Marit Westergaard. 2018. The acquisition of word order in L2 Norwegian: The case of subject and object shift. *Nordic Journal of Linguistics* 41(3). 247–274. https://doi.org/10.1017/S0332586518000203.

Bardel, Camilla & Laura Sánchez. 2017. The L2 status factor hypothesis revisited. In Tanja Angelovska & Angela Hahn (eds.), *L3 Syntactic Transfer: Models, New Developments and Implications*, 85–101. Amsterdam: John Benjamins.

Bardel, Camilla & Ylva Falk. 2007. The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research* 23(4). 459–484. https://doi.org/10.1177/0267658307080557.

Beslagic, Deni. 2024. Minding the English trap: A crosslinguistic pedagogical approach to teaching the V2 rule to young German L3 learners in Sweden. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 29(2). 137–161. https://doi.org/10.48694/zif.3854.

Bohnacker, Ute. 2006. When Swedes begin to learn German: From V2 to V2. *Second language research* 22(4). 443–486. https://doi.org/10.1191/0267658306sr275oa.

Bohnacker, Ute & Marit Westergaard. 2024. Learning and unlearning Verb Second word order. In Tania Ionin, Sivina Montrul & Roumyana Slabakova (eds.). *The Routledge Handbook of Second Language Acquisition*: *Morphosyntax and Semantics*, 355–367. New York: Routledge.

Busterud, Guro, Anne Dahl, Dave Kush & Kjersti F. Listhaug. 2023. Verb placement in L3 French and L3 German: The role of language-internal factors in determining cross-linguistic influence from prior languages. *Linguistic Approaches to Bilingualism* 13(5). 693–716. https://doi.org/10.1075/lab.22058.bus.

Clahsen, Harald. 1988. Critical phases of grammar development. A study of the acquisition of negation in children and adults. *Language development* 5. 123–148. https://doi.org/10.1515/9783110854817-010.

Dahl, Anne, Kjersti F. Listhaug & Guro Busterud. 2022. The role of L1 Norwegian and L2 English in the acquisition of verb placement in L3 German. In Tania Leal, Elena Shimanskaya & Casilde A. Isabelli (eds.), *Generative SLA in the age of minimalism*, 191–212. Amsterdam: John Benjamins.

Duden online. 2025. Häufigkeit. https://www.duden.de/hilfe/haeufigkeit, 29 July 2025.

DWDS [Digitales Wörterbuch der deutschen Sprache]. 2025. 'jetzt'. https://www.dwds.de/wb/jetzt, 29 July 2025.

Falk, Ylva & Camilla Bardel. 2011. Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research* 27(1). 59–82. https://doi.org/10.1177/0267658310386647.

Falk, Ylva & Christina Lindqvist. 2022. Teachers' attitudes towards multilingualism in the foreign language classroom: The case of French and German in the Swedish context. In Anna Krulatz, Georgios Neokleous & Anne Dahl (eds), *Theoretical and applied perspectives on teaching foreign languages in multilingual settings: Pedagogical Implications 100*. 154–169. Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781788926423-014

Flynn, Suzanne, Claire Foley & Inna Vinnitskaya. 2004. The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism* 1(1). 3–16. https://doi.org/10.1080/14790710408668175.

Garbacz, Piotr & Janne Bondi Johannessen. 2014. Negation in imperative clauses. *Nordic Atlas of Language Structures (NALS) Journal 1*, 250–265.

Geher, Glenn & Sara Hall. 2014. *Straightforward statistics: Understanding the tools of research*. Oxford University Press.

Han, Chung-hye. 2001. Force, negation and imperatives. *The Linguistic Review* 18. 289–325. https://doi.org/10.1515/tlir.2001.004.

Haukås, Åsta. 2016. Teachers' beliefs about multilingualism and a multilingual pedagogical approach. *International journal of multilingualism* 13(1). 1–18. https://doi.org/10.1080/14790718.2015.1041960.

Hawkins, Roger. 2001. *Second language syntax: A generative introduction*. Oxford: Blackwell.

Helbig, Gerhard & Joachim Buscha. 2000. *Leitfaden der deutschen Grammatik*. Berlin & München: Langenscheidt.

Hermas, Abdelkader. 2014. Multilingual transfer: L1 morphosyntax in L3 English. *International Journal of Language Studies* 8(2). 1–24.

Jensen, Isabel Nadine, Natalia Mitrofanova, Merete Anderssen, Yulia Rodina, Roumyana Slabakova & Marit Westergaard. 2023. Crosslinguistic influence in L3 acquisition across linguistic modules. *International Journal of Multilingualism* 20(3). 717–734. https://doi.org/10.1080/14790718.2021.1985127.

Kim, Hae-Young. 2017. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative dentistry & endodontics* 42(2). 152–155. https://doi.org/10.5395/rde.2017.42.2.152.

Mackey, Alison & Susan M. Gass. 2015. *Second language research: Methodology and design*. London: Routledge.

Meisel, Jürgen M. 1997. The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second language research* 13(3). 227–263. https://doi.org/10.1191/026765897666180760.

Nettskjema. 2024. Nettskjema. https://nettskjema.no/?lang=en, 24 April 2024.

Pittner, Karin & Judith Berman. 2010. *Deutsche Syntax: Ein Arbeitsbuch*. Tübingen: Narr Francke Attempto Verlag.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry* 20(3). 365–424. https://www.jstor.org/stable/4178634.

Puig-Mayenco, Eloi, Jorge González Alonso & Jason Rothman. 2020. A systematic review of transfer studies in third language acquisition. *Second Language Research* 36(1). 31–64. https://doi.org/10.1177/0267658318809147.

Rindal, Ulrikke. 2019. PhD revisited: Meaning in English: L2 attitudes, choices and pronunciation in Norway. In Ulrikke Rindal & Lisbeth M. Brevik (eds.): *English Didactics in Norway: –30 years of doctoral research*, 335–355. Oslo: Universitetsforlaget.

Rothman, Jason. 2015. Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: language and cognition* 18(2). 179–190. https://doi.org/10.1017/S136672891300059X.

Slabakova, Roumyana. 2017. The scalpel model of third language acquisition. *International Journal of Bilingualism* 21(6). 651–665. https://doi.org/10.1177/1367006916655413.

Stadt, Rosalinde, Aafke Hulk & Petra Sleeman. 2018. The influence of L1 Dutch and L2 English on L3 French: A longitudinal study. *Journal of the European Second Language Association* 2(1). 63–71. https://doi.org/10.22599/jesla.42.

Stadt, Rosalinde, Aafke Hulk & Petra Sleeman. 2020a. L2 influence in L3 acquisition: The role of the L3. In Andreas Trotzke & Tanja Kupisch (eds.), *Formal Linguistics and Language Education: New Empirical Perspectives*, 203–221. Cham: Springer.

Stadt, Rosalinde, Aafke Hulk & Petra Sleeman. 2020b. L1 Dutch vs L2 English and the initial stages of L3 French acquisition. In Camilla Bardel & Laura Sánchez (eds.), *Third language acquisition: Age, proficiency and multilingualism*, 237–261. Eurosla Studies. Berlin: Language Science Press.

Westergaard, Marit. 2003. Unlearning V2: Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA yearbook* 3(1). 77–101. https://doi.org/10.1075/eurosla.3.07wes.

Westergaard, Marit. 2021a. Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research* 37(3). 379–407. https://doi.org/10.1177/0267658319884116.

Westergaard, Marit. 2021b. L3 acquisition and crosslinguistic influence as co-activation: Response to commentaries on the keynote 'Microvariation in multilingual situations: The importance of property-by-property acquisition'. *Second Language Research* 37(3). 501–518. https://doi.org/10.1177/02676583211007897.

Westergaard, Marit, Natalia Mitrofanova, Roksolana Mykhaylyk & Yulia Rodina. 2017. Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism* 21(6). 666–682. https://doi.org/10.1177/1367006916648859.

Westergaard, Marit, Natalia Mitrofanova, Yulia Rodina & Roumyana Slabakova. 2023. Full Transfer Potential in L3/Ln acquisition: Crosslinguistic influence as a property-byproperty process. In Jennifer Cabrelli, Adel Chaouch-Orozco, Jorge González Alonso, Sergio Miguel Pereira Soares, Eloi Puig-Mayenco & Jason Rothman (eds.), *The Cambridge Handbook of Third Language Acquisition and Processing*, 219–242. Cambridge: Cambridge University Press.

Wojtecka, Magdalena, Corinna Koch, Angela Grimm & Petra Schulz. 2011. Production and comprehension of sentence negation in child German. In Angela Grimm, Anja Müller, Cornelia Hamann & Esther Ruigendijk (eds.), Production-Comprehension Asymmetries in Child Language (S. 217–246). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110259179.217.

ZS ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 5
https://doi.org/10.18148/zs/2026-2005

Kai-Uwe Carstensen, University of Siegen, Germany, Kai-Uwe.Carstensen@Uni-Siegen.de

When "come" means "get"

On the semantics of directional deictics

Abstract: Similar to locative deictics like *here* and *there*, directional deictics like *come*, *go*, *hither*, *thither* are hard to characterize semantically. Inspired by the seminal work of Charles Fillmore on spatial deixis, the alleged movement-based contrast of *come* and *go* has been in the fore of interest for some time. In the present article, it will be shown that this focus has to be widened to include other spatial verbs (and the adverbs/particles), and that the semantic criteria have to be specified more abstractly to encompass non-actual locomotion as well as non-spatial aspects of objective and subjective valuation, and to explain unexpected patterns of (un)acceptabilities. Based on data from different languages (especially taking the peculiarities of German into account), it will be argued that the deictic oppositions can be attributed to attention-based semantic criteria. These specify the perspective of directional deictics as focus on one of the poles of a state change, with the salience of the conceptually instantiated states being determined by criteria of relevance-based or qualitative valuation in the actual context of language production.

Keywords: Attention, cognitivist semantics, directional deictics, deixis, spatial semantics

Cite as: Carstensen, Kai-Uwe. 2026. When "come" means "get": On the semantics of directional deictics. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi.org/10.18148/zs/2026-2005
Received: 06 May 2025; **accepted:** 30 November 2025; **published:** 30 June 2026
Published by: publikon / dgfs (Linguistic Society of Germany)

1 Introduction

Spatial deixis, the field of highly context-dependent spatial expressions (SPATIAL DEICTICS), can be subclassified into locational deixis (including expressions like *here* and *there*, see Carstensen 2023) and DIRECTIONAL DEIXIS (*come, go, hither, thither* etc.), the topic of the present article. Spatial deictics are often associated with the concept of 'deictic center' (for overviews see, e.g., Oshima 2006a, Nakazawa 2007) or the notion 'origo' by Bühler, see Bühler 1990 [1934]), including the notion of directly pointing/referring to elements of the actual context of some utterance.[1]

In his studies of the semantics of *come* and *go* (and their causatives *bring* and *take*, respectively, see Hockett 1990) Fillmore (e.g., Fillmore 1975) departed from analyses in which *come* and *go* are assumed to indicate movement toward or away from the deictic center, respectively. Instead, he tried to specify the conditions that hold for the appropriateness of the verbs' use in some language (Oshima 2006a, 2006b, Nakazawa 2007).

Fillmore characterizes the verbs in question as DEICTIC VERBS OF (LOCO)MOTION, where "locomotion" (also called "translational motion") is defined as "an object changing its location through time" (1975:25), usually associated with continuous movement. This suggests, despite finer distinctions discussed, a complementary opposition of *come* and *go*, ('locomotion toward reference point' vs. 'locomotion not toward reference point'), which some have taken to be (near-)universals (or primes) of reference point related movement (COME/GO,[2] see the Natural Semantic Metalanguage of Wierzbicka critically assessed in Goddard 1997).

However, such a view has been refuted in a detailed crosslinguistic study with Australian and Austronesian languages (Longgu, Mparntwe Arrernte) by Wilkins & Hill (1995), which showed that crosslinguistic variance of corresponding verbs does not allow a generalization of COME and GO as universals and that there are languages in which the "go" verb is not inherently deictic.[3] Likewise, Bohnemeyer (2010) found that translational motion is not necessarily coded linguistically by "movement verbs". For Yucatec Maya he showed that actual movement of a FIGURE/theme is rather expressed by "verbs of change of location", which can even be used in cases where there is no motion of the figure at all (instead, the CHANGE OF LOCATION is caused by movement of the GROUND/relatum). In general, there is some typological diversity in the syntactic/morphological/lexical realization of VENTIVE (~"coming" expressions) and ITIVE (~"going" expressions) constructions, and their relationship to some directed (bounded) movement (CENT-

[1] But see the criticism of these approaches and a more elaborate outline of (spatial) deixis in Carstensen (2023).
[2] Capitalized English terms are typically used to refer to (hypothetical) non-linguistic, conceptual units that may be expressed differently cross-linguistically. Double quotation of a lexeme is used to easily refer to such expressions: for example, "come" verbs as verbs meaning 'COME' (*come, kommen, venir,...*).
[3] This a selective summarization of their much more elaborate analysis.

RIPETAL 'to speaker', CENTRIFUGAL 'away from speaker', TRANSVERSE 'passing across') is not at all clear-cut (see Fortis & Fagard 2010 for an overview).

Furthermore, some have raised doubt that the "come"/"go" contrast is exclusively associated with a spatial context. Clark (1974) observes that non-literal or idiomatic uses of *come, go, bring, take* and *send* do not involve motion, but abstract changes with respect to objective or subjective dimensions of valuation. Similarly, Kuno & Kaburaki (1977) investigate cases of linguistic contrast based on EMPATHY prominently found in Japanese and point to their relevance for deixis.

The present paper explores the semantics of directional deictics in the light of these controversies. With a review of the semantic analyses of "come"/"go" verbs in Section 2, it will be shown that current approaches do not fully account for the range of observed phenomena, and that the scope of data is too restricted for an adequate analysis. By discussing (further) data from languages as close as English and German in Section 3, it is argued that they not only exhibit all the mentioned problems but also provide hints for a theoretical solution. The article takes up a cognitivist position on this topic and approaches directional deixis from the question of which representations and processes are involved in its systemic structure and actual genesis. Central aspects of this are presented in Section 4, including SELECTIVE ATTENTION playing a central role for directional deixis, and the LANGUAGE GENERATION VIEW of the approach. A third aspect concerns the role of attention-based PERSPECTIVATION in a two-level semantics, given the constraints posed by the variance in the data and the theoretical need to both find generalizations ("universals") and account for specific differences. Section 5 collects and succinctly presents a number of topics relevant in the semantics of directional deictics, and Section 6 shows how these are reflected in a proposal for the latter – as a specific, but representative semantic analysis of selected directional deictics that accounts for the discussed perspectivation problems.

2 On "coming" and "going"

(1) summarizes what Oshima (2006b: 109) calls the "classical analysis" of *come* and *go*. It involves the notion of the speaker as the DEICTIC CENTER (Bühler´s origo of a multidimensional coordinate system) of the speaking context, and of PERSPECTIVE (perspectivation).

(1) [from Oshima (2006b: 109)]
   a. *Go* describes motion *from* the speaker or his proxy (the individual whose perspective the speaker takes).
   b. *Come* describes motion *to(ward)* the speaker or his proxy (the individual whose perspective the speaker takes).

Fillmore notes that the interpretation of deictic expressions may involve more than reference to the actual context: “For example, if I say, ‘I came there yesterday morning’, it cannot be that I am there now, because ‘there’ is by definition a place where I am not now located, and it cannot be that I was already there yesterday morning when I came” (Fillmore 1975:10). To allow for other elements to constitute the corresponding center for a description (addressee, referent in (2)), ‘deictic center’ models therefore have to either abstract from the actual linguistic/coding context (hence “proxy”) or work with shifts of the deictic center (or even with different kinds of origos, cf. Carstensen 2023).

(2) [from Oshima (2006b: 110)]
   a. *Can I come visit you?*
   b. *John was preparing a meal. Then, the cat came to him.*

According to Oshima, ‘deictic center’ models suffer from at least two problems. The first appears in the pattern of acceptability in (3), which runs counter to expectations: shifting to the addressee is preferred in (3a), but is excluded in (3b). The second concerns the unexplained possibility of multiple elements to fill the role of the deictic center at the same time, given that perspective is theoretically not expected to vary within a sentence ((4)).

(3) [after Oshima (2006b: 110f)]
   a. *Can I ??go/come visit you?*
   b. *Will you *go/come visit me?*

(4) [after Oshima (2006b: 111)]
   a. *At least two students {went/came} to talk to three professors.*
   b. *He’ll come to your house before he comes to my house.*

An even harder problem for ‘deictic center’ models (but also others) is the differing acceptability in (5). A similar example is (6), which Fillmore presents as rejected by many (compare his (4b)). Overall convinced that “[t]he hypothesis that there might be only one deictic center in conversational discourse got disconfirmed”, he regards this as a “horror” and has “no idea on earth what to say about it” (Fillmore 1975: 68).

(5) [from Kuno & Kaburaki 1977: 664]
   a. *I went to the party and found that John was coming to the party, too.*
   b. **I came to the party and found that John was going to the party, too.*

(6) *??He’ll come to my house after he comes to your house.*

Instead of utilizing the wholistic concept of ‘deictic center’, Fillmore identifies a number of factors determining the appropriateness of using *come/go*, especially for the goal condition of *come*. Basically, they comprise aspects of the linguistic act: person (speaker, addressee, other),

place (person's location), and time (utterance/"coding" time, event/"reference" time). In non-actual, narrative conditions, person and place can be instantiated by the central character or place of the narrative, and the goal may only be the *home base*[4] of the person. Also, the company of a relevant person was found to be a factor.

However, it can be observed that the distinction of *come* and *go* is not as concise as probably desired. This is also criticized by Oshima who discusses examples with an unexpected preference for *come* where there is motion toward a person which is neither the speaker, nor the addressee, nor even a unique central character of a narrative ((7),(4a)). As an alternative, Oshima introduces the concept RP ("Reference Point") referring to a set of individuals, and proposes the corresponding conditions listed in (8). Membership in the RP is then supposed to be subject to constraints that are language-specific and context-dependent to some extent, for example (i) the speaker is always member of the RP, (ii) the addressee is preferred to be a member, (iii) a third person may be a member if it is discourse-salient.

(7) *It seems that John came to my sister's place yesterday.*

(8) [from Oshima (2006b: 114)]
   a. *Go* requires that no member of the (contextually determined) RP be at the goal of the described motion at the utterance time.
   b. *Come* requires that (i) there be some member of the RP at the goal of the described motion at the utterance time or at the event time, or (ii) the goal be the home base of a member of the RP (at the event time).

Formally, Oshima separates the assertional parts of *come* and *go* (movement of a theme to some goal) from the presuppositional parts (no member of RP at goal for *go*, some member of RP at goal for *come*), and treats the RP as a context parameter (like time, place, speaker etc.).[5] Noting that variable RP sets as context parameters would be an unconventional solution for the problem, Barlew (2017) rather treats members of the RP as "anchors" (i.e., elements with salient perspectives in discourse[6]) that behave like anaphors, which he motivates with quantificational and more complex examples like (4a) and (9). With this, Barlew explicitly analyzes the shift in deictic perspective (from speaker/addressee to others) via/as anaphora resolution. This approach does not account for the acceptability differences of (4b) and (6), however. Also, in Fillmore's example (10) (after Levinson 2006: 103), the different perspectives cannot be explained

[4] As Goddard (1997: 157) points out, "home base" does not need to refer to the *actual* home (base) of some person, as in "It's a pity John's coming to the shop tomorrow, when neither of us will be there".
[5] According to Sudo, only *come* has a context-dependent presupposition (roughly, 'home location of RP is goal', see Sudo 2018).
[6] Which means that the anchor is a relevant "perspective holder" (Anderson 2021) but does not actually need to be at the goal: "Ron thinks he's in New York, and he thinks Tom is coming to New York" (both interlocutors in Columbus, Barlew 2017: 42).

with RP or corresponding anchors. Barlew recognizes the problem with RP-less situations and offers the concept 'perspective-only anchoring', according to which the anchor's perspective is centered at the destination for *come*, even if the anchor is not located there.

(9) [Context: Ann is in Cleveland, Ben is in New York, and Chris is in Denver. On the phone, Ann asks Ben *Where is John these days?*. Ben says:]
Chris {thinks/says} that John is coming to Denver today.

(10) *The door of Harry's lunchroom opened and two men {came/went} in.*

Data from crosslinguistic/typological research present a vast range of variance in directional deixis (e.g., Wälchli & Zúñiga 2006). Longgu (Wilkins & Hill 1995), for example, has a non-deictic word *la* 'go', which in combination with *mai* 'hither' is an expression for *come*, and with *hou* means 'go thither/from here'). Yet while – in the specific experimental setting used – Longgu "come" and "go" always relate to the deictic center (movement starts/ends at the (location of the) speaker), Mparntwe Arrernte is less strict in that respect ("come" may also have a "toward" meaning). Palauan (see Nakazawa 2007: 63) has two verbs for *come*, depending on whether movement is toward the speaker (*me/mei*) or hearer (*eko*), else it is *mo/mong* (except for auxiliar *mo*, which means 'become, get'). Interestingly, not all languages with "come" verbs have the same freedom in choosing reference points other than the speaker. In Japanese, perspective shift to the addressee is only possible if the moving figure is not the speaker (Nakazawa 2007: 66). While in Chinese the reference point for "come" is restricted to the speaker regardless of time, it is even restricted to utterance time in Shibe (Nakazawa 2007: 78f).

With her investigation of non-literal or idiomatic uses of "come/go/bring/take/send" verbs, Clark (1974) extends the range of phenomena in another direction. She observes that the verbs in examples like (11) do not involve locomotion, but abstract changes with respect to objective or subjective dimensions of valuation. According to her, *come* implies TRANSITION TO A "NORMAL STATE" ((11a), (11b)) or to a state that is evaluated as positive by the speaker ((11c)). Interestingly, this is slightly different for *go*, because – apart from denoting departure from a normal state in normal state deixis – in the evaluative deixis of (11c), "the speaker is quite neutral about the protagonist's destination, and the focus of the utterance seems to be on the ordeal rather than on its ending" (Clark 1974: 366f). As Malsch & Lant (1977) observe, however, Clark's "normal state deixis" is violated by cases in which things are described as *coming unraveled, unglued, or apart* (*or undone*).

(11) [after Clark (1974)]
a. *The motor {went/*came} dead.*
b. *The motor {came/*went} {alive/to life} again.*
c. *Look at all he {came/went} through.*

An abstract asymmetry overlaying a spatial change of location thereby establishing different points of view is what Kuno & Kaburaki (1977) find in Japanese. Here, a speaker has to choose one of two Japanese words for *give* depending on whether the agent (~source) has a higher subjective EMPATHY VALUE than the recipient (~goal) or not (all else being equal). They also state: “Undoubtedly, the choice between *come* and *go* in English and many other languages is determined by empathy factors” (Kuno & Kaburaki 1977: 663). Yet they note an unexpected difference in acceptability in multi-clause examples like (12), for which they have no explanation. These examples correspond to those problematic for Fillmore (see (4b) and (6)).

(12) [from Kuno & Kaburaki (1977)]
  a. *I will not go to a party that John is coming to.*
  b. **I will not come to a party that John is going to.*

3 Beyond “come” and “go”: A closer look at German (and English)

In (13), the literary version of the famous grave inscription at Thermopylae by Friedrich Schiller, the author uses *kommen* ‘come’ in addressing the wanderer. In this case, both the implicit speaker and the addressee are far from the goal (which is also not a home base), and the movement is explicitly hypothetical (note also the use of *there*). (14) is part of the traditional song about Hänsel and Gretel (from Grimms’ fairy tales) and describes their arrival at the witch’s house. Since the latter has not been introduced so far and therefore does not count as an RP location or known home base, the justification for the use of *kommen* remains unclear as well.

(13) *Wanderer, kommst du nach Sparta, verkündige dorten, du habest uns hier liegen gesehn, wie das Gesetz es befahl.*[7] (German)
‘Wanderer, if you come to Sparta, proclaim there that you saw us lying here, as law has commanded us.’

(14) *Sie kamen an ein Häuschen [...]* (German)
‘They came to a little house [...]’

Similarly, in (15), both speaker and addressee are definitely not in the goal location (which is the very problem of A and B!). Yet while the use of *come* is unacceptable in English, its German equivalent is perfect.

(15) [A and B want to enter a cave, the iron door is locked][8]

[7] Friedrich Schiller, “Der Spaziergang”, 97f (1795, on https://de.wikisource.org/wiki/Der_Spaziergang_(Friedrich_Schiller), acc. 2025-10-24).

a. *How do we *come/*go/get in there?*
b. *Wie kommen/*gehen/gelangen wir da rein?* (German)
‘How do we come/*go/get in there?’

(15) also demonstrates the basic relevance of a third spatial verb (*gelangen*/*get*) as a natural opposite of the “come/go” verbs, usually not considered in this discussion. Altogether, the use of the verbs here shows a pattern of acceptability that is in conflict with standard assumptions. As noted, the “come” verb is unexpectedly acceptable in German, which would require a further generalization of the reference point conditions. But then, as English *come* seems to obey the rules, *go* should also be perfect, which is not the case. Technically, Star Trek and beaming being science fiction, entering the cave requires movement. Hence, one of the “come/go” forms (or even both) should be adequate in this situation. From this it follows that the scene is conceptualized and linguistically framed as a CHANGE OF LOCATION in these non-exotic languages, and that *gelangen (get)* (but also *kommen* in this sense) are the corresponding verbs. Such a view is supported by the contrast of **Get!/*Gelang!* with *Come!/Go!*, pointing to the role of a semantic movement/action component: it can be emphasized/focused in the case of movement verbs (leaving the omitted path component to pragmatic interpretation), but is missing (by definition) in the semantics of change-of-location verbs.

Slightly different from the *come/go* contrast, the use of *come/get* is clearly a matter of perspective: in (15), *gelangen (get)* is the lexically contrasting term to *kommen*, similar to the lexical contrast of *come* and *get* in (16). Yet *gelangen* is not a casual term and has positive expressive meaning, which is the reason why *kommen* is preferred in (15) (*gelangen* cannot even be used in a translation of (16b)). A more negative quality of the goal can be expressed by the verb *geraten (get)*.

(16) [A in room to B outside]
a. *Come in!*
b. *Get in here, now!*

Similar observations can be made with respect to BRING and TAKE. They are usually portrayed as causative versions of COME and GO (Fillmore 1975, Preston 1984), which is most explicit in Yucatec with its causal suffix, compare *bin* “go”/ *bis* “take” and *tàal* “come” / *tàas* “bring” (Bohnemeyer 2010: 115). As (17) shows, there seems to be a different lexeme (*holen/get*) as (further) opposite term of the “bring” verbs. However, English and German differ in that German *bringen* does not express BRING, but more general TRANSPORT, as in (18). Accordingly,

[8] A discussion of similar examples, albeit (only) from a restricted, typological viewpoint, can be found in: Davide Ricca, *I verbi deittici di movimento in Europa: una ricerca interlinguistica* (Firenze: La Nuova Italia, 1993).

holen is the corresponding term for *bring* (actually, for *fetch*) and *schaffen* is a low-register lexical alternative for *bringen (get).*

(17) [A does not have the file]
  a. A: *Bring/*take/?fetch*[9]*/get me the file!*
  b. A: *Bring/*nimm/hol mir die Akte!* (German)

(18) [A and B are in the house]
  a. A: *Bring/*nimm/schaff/*hol den Müll raus!* (German)
  b. A: *Take/get out the garbage!*

Once adverbs or particles of direction are in theoretical focus, a problem appears in German: it has three-component directional specifications (verb + locational deictic adverb + directional particle) as in the prototypical cases (19) and (20) (different from English, verb + locative is ungrammatical as a directional).[10] This necessitates an explanation for the fact that – different from Longgu – only selected combinations are acceptable (see (21)).[11] Besides that, German has particles/adverbs that systematically combine directional and locational information ('hither + in' -> *herein*, 'thither + in' -> *hinein*). Although *hereinkommen* and *hineingehen* are the expected prototypical complex constructions, however, particle acceptability is different in the cave scenario of (15), see (22).[12]

(19) *Komm hier her!*
come here-ADV hither-PART
'come here/hither'

(20) *Geh da hin!*
go there-ADV thither-PART
'go there/thither'

(21) [German three-component deixis combinatorics]
  a. *Komm hier hin! / Komm da hin!*
  b. *?Geh hier her! / ?Geh hier hin!*[13]
  c. **Komm da her! / *Geh da her!*
  d. *Komm her! / *Komm hin!*

[9] Preston (1984: 185) advises English learners to "[s]coff (except in dog-training) at any *fetch*" for BRING.
[10] Although German *hin* roughly corresponds to the English expression of 'from here' and *her* to that of 'to here', *hin* and *her* will be translated by *thither* and *hither*, respectively, in the following.
[11] It is important to point out that acceptability ratings can be different in some dialect. For example, Mosel Franconian is known for its use of *hergehen* 'go hither', and in some parts for systematically using *holen* 'get' for *nehmen* 'take'.
[12] While *rein* could be assumed to be an abbreviated *herein*, it is rather a generalized deictic directional adverbial *in* 'into' where the RP can be associated with source or goal, depending on the verb. This is not an idiosyncrasy, as shown by *Geh/komm raus/rüber/runter!* 'Go/Come out/over/down!'.

(22) [A and B want to enter a cave, the iron door is locked]
*Wie kommen wir da rein/*herein/hinein?*

German is furthermore quite different from English in its non-literal use of directional deictics. Rauh (1981) shows that German directional deixis metaphors do not express normal state or evaluative asymmetries, as someone can "kommen" 'come' to hell or heaven, depending on her behaving, and a negative adjective (*teuer* 'expensive') is used with that verb in the phrase *das wird dich teuer zu stehen kommen* 'that will cost you dearly'. Qualitative changes are rather uniformly expressed by *werden* 'become/get'. If at all, *kommen* is used to focus on the final state (*zu Schaden kommen/*gehen* 'get ruined'), while *gehen* emphasizes the cause or action (*(*Komm/) Geh mir aus den Augen* 'Leave me alone', lit. "Go out of my eyes"). Overall, this section has shown that data from German and English not only show certain phenomena found in exotic languages (e.g., change-of-location use of directional verbs), but also –although quite close– many interesting differences to be treated in semantic theories of directional deictics.

4 Aspects of Cognitivist semantics

Gerwien & von Stutterheim (2022b) argue that it is necessary to extend the scope of investigation in language and directional space from the linguistic format across languages to the cognitive processes that determine the construal of motion events and subsequent encoding. In a review of this research field based on comparative studies with psycholinguistic and neuroscientific experiments, they show that there is a close coupling of perceptual and linguistic processing that challenges the assumption of a clear language/thought divide and points to language-specific influences on actual cognitive perspective-taking. For example, their own work with Tunisian Arabic speakers (Gerwien & von Stutterheim 2022a) reveals that different grammatical coding of manner, place and aspect of L1 Tunisian Arabic impacts event descriptions in L2 German compared to native German speakers (aspectual differences are also reflected in different patterns of visual attention to the described scenes as measured by event related potentials of German and English speakers, see Flecken et al. 2015). In their language production oriented cognitive model, language-specific selection of information for speaking is stored as so-called "pragmatic experience" in the form of linguistically biased event schemata that guide the description of new motion scenes. By describing these schemata as "attentional fil-

[13] This is awkward in standard German. If I am unsure here, it might derive from the fact that German *gehen* has GO and WALK senses and that the sentences might express a focus on the specific action to be performed. In questions, the combination is clearly out: **Woher gehst du?* 'Whence do you go?'. Note that other acceptable uses of *hin/her* are usually not deictic, but only express directionality, for example *hin und her* 'to and fro', or accompaniment, for example *hinter/vor ihm her gehen* 'go behind/in front of (i.e., follow or lead) him'. These aspects are also found in regional use of abstract action senses of *hergehen* (*da gehen wir her und …* 'we act as follows: …'), as well as in the corresponding standard German noun *Hergang* 'course of events/action (e.g., of a crime)'.

ters at the interface between perception and selective representation, as well as at the interface between information selection from a knowledge base and verbal encoding" (Gerwien & von Stutterheim 2022a: 12), the authors highlight the relevance of selective attention and language production. Both are also central to the cognitivist approach put forward here, as is a general attention-based two-level approach to semantics. To explain their role for the semantics of directional deictics, these three aspects are discussed in this section.

4.1 Selective attention

Although "[e]veryone knows what attention is" (James 1890: 404),[14] most people are surprised when encountering so-called *inattentional blindness*, a condition in which we do not register visible things – even a gorilla walking plainly before our eyes – because attention is distracted by some task (see Simons & Chabris 1999). (Neglect) phenomena like these point to the fact that not everything we see is available for thinking, speaking or action, and that it requires SELECTIVE ATTENTION to make elements of some input level representation available for further processing. This has led to computational models (e.g., Itti & Koch 2001) in which selective attention plays a central role: it mainly operates bottom-up/data-driven based on input, but can also be influenced/modulated top-down/knowledge-driven.

Cognitive representations of the world do not passively reflect a somehow given structured world, but are cognitively *constructed* based on the partial perceptual information about what is out there and on evolutionary successful representations of physical regularities. (Loco)Motion is a case in point.[15] While there is certainly continuity in the physical movement of things changing location, (LOCO)MOTION is just an ascription based on the interpretation of successive perceptual "snapshots". This constructive aspect can be illustrated with visual illusions of *apparent motion* (Ramachandran & Anstis 1986) arising in displays that lack actual continuity. In one of their experiments, Ramachandran & Anstis show subjects continually alternating pairs of diagonal dots ("bistable (dual state) quartet", see Figure 1). What the observers report (under some spatiotemporal conditions of the display) is continuous motion of the dots, either streaming horizontally to and fro (a), or bouncing up and down (b). Eventually, circular movement of a dot can be observed (clockwise or counter-clockwise) and sometimes even intentionally induced by attentional modulation (Kohler et al. 2008).

[14] What I actually mean is this: "I do not like the word attention – it has too many colloquial connotations – but I don't know of a better label" (Graziano 2020: 228).

[15] "We believe motion is controlled in the early stage of visual processing by what is in effect a bag of tricks, one the human visual system has acquired through natural selection during millions of years of evolution" (Ramachandran & Anstis 1986: 102).

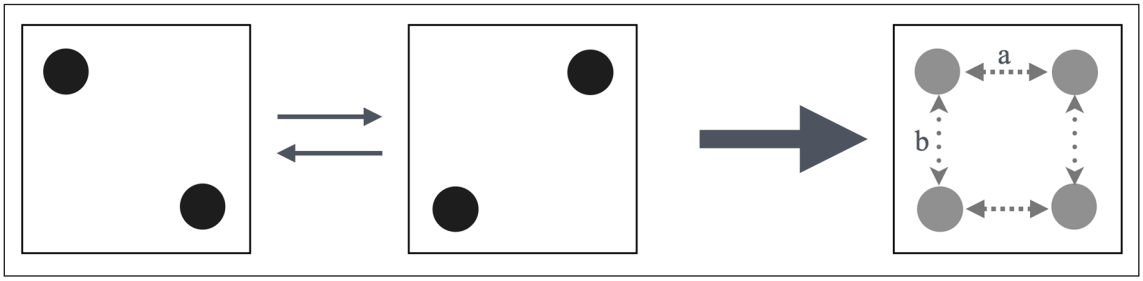

Figure 1: Apparent motion from Bistable Quartet

Beyond aspects of cognitive *processing*, selective attention also has a *representational* constructive side as shown in Carstensen (2011): the basic structure of non-linguistic conceptual representations as a whole can be regarded as a system of represented attentional engagement (see also Graziano 2020 for a proposal of "attention schemata" as higher order representations of attentional processing), where the upper level of such a "Cognitivist ontology" is defined by basic distinctions (global/local processing, focused/distributed attention) relative to temporally-delimited contexts ("frames") within some domain. The *sequence* of attentional engagement (continuous ATTENTIONAL SCANNING or single ATTENTIONAL CHANGES) establishes higher-order representational elements on some level and in some domain, for example spatial relations (Carstensen 2015), paths (Carstensen 2019), or grade relations (Carstensen 2013).

Attentional criteria are therefore definitory for conceptual entities and determine the structure of non-linguistic, non-perceptual representations. Carstensen (2011) discusses the upper structure of "objectual-level" (objects and stuff) and "eventuality-level" (events and states) entities as basic conceptual distinctions (the latter depicted in Figure 2, adapted from and explained in Carstensen 2011). The eventuality-level categories obviously correspond to the familiar Vendlerian categories of *lexical aspect* (or *Aktionsart*) (compare, e.g., *stand*, *go*, *enter*, *pass*).

As to spatial representations, the same attentional criteria lead to spatial-level, mostly directional, distinctions homologous to the eventuality-level ones (see Figure 3 adapted from and explained in Carstensen 2019) (compare, e.g., *on*, *through*, *into*, *past*). This similarity in structure is rarely found in the literature. Typically, conceptual structure and spatial structure are viewed as juxtaposed systems (see Jackendoff 2019), compositionally connected via path information for directionals. In that case, it is a challenge to ensure aspectual congruence of verb and spatial prepositional phrase (PP) (*stroll through the woods*/**into the house*, "*enter the house through the garden"). Viewing spatial-level and eventuality-level representations as continuous on a perception-to-cognition gradient, and linked by a dedicated operator in semantics, offers an explanatory solution for this problem as shown in Carstensen (2019).

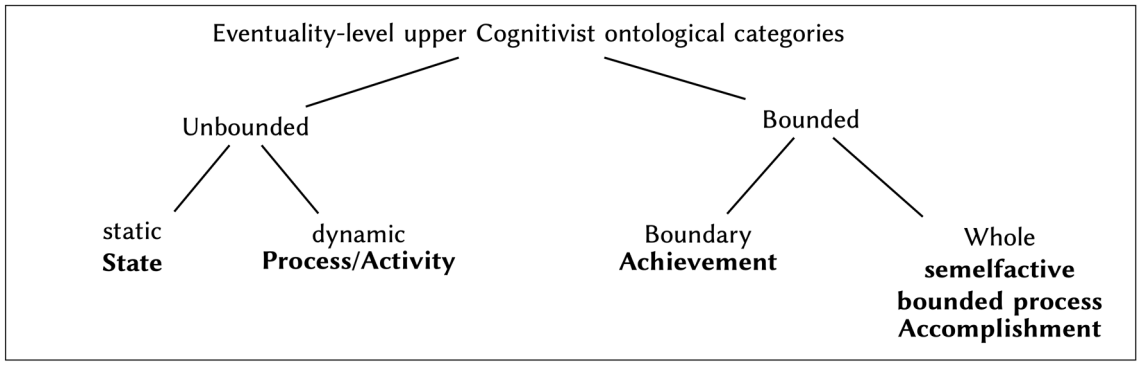

Figure 2: Eventuality-level upper ontology. Adapted from Carstensen (2011)

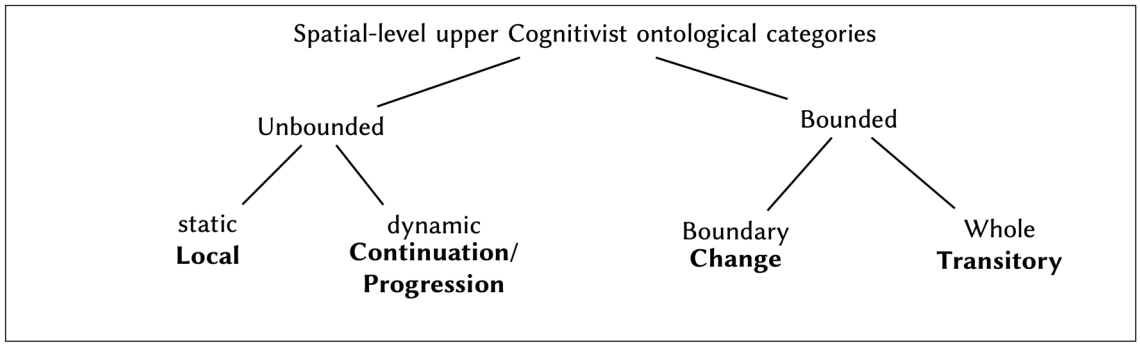

Figure 3: Spatial-level upper ontology. Adapted from Carstensen (2019)

4.2 Language Production

Work in semantics – especially in formal approaches – mostly adopts the interpretation perspective, and the importance of *language generation/production* and the *bidirectionality* of semantics as the interface between syntax and conceptual representation is rarely put to the fore (see also Carstensen 2021 on this point). Typically, the theoretical focus is only on what an expression (objectively) *means* or *denotes*, as opposed to on what the conditions are (i.e., aspects of structure and processing) to *select* it, given some situation and context. And yet, a language production perspective is highly relevant and somehow primary: before a word is interpreted, it must have been produced by someone, including lexical choices based on the semantic framing of some conceptualized perceptual content in the given context. As Jackendoff (2019: 110) writes: "one can now come to view language not as a system that *derives* meanings from sounds (say prooftheoretically), but rather as a system that *expresses* meanings" (his emphasis). This usage-oriented view of semantics is most clearly expressed by Levelt: "A lexical item's lemma information consists of the conceptual specification *for its use (including pragmatic and stylistic conditions)*[…]" (Levelt 1989: 223, my emphasis).

In language production, one encounters the so-called "hyperonym problem" (Levelt 1989: 201): if we look at, say, a picture of a parrot (having the typical features of a bird), why don't we always select *bird* although parrots are less frequent than (all) birds, and are instances of a more specific type than 'BIRD', and although the selection of *bird* should be much easier and quicker from a processing point of view (spreading activation etc.)? Mitchell et al. (2013) pointed out that language production is less based on the contrastiveness of features (and therefore prone to over- or underspecification) and much more coupled to actual perception (which mirrors the insights gained in language understanding, see Tanenhaus et al. 1995). For the simple case of generating referring expressions for objects in the visual domain they refer to *visual attention* having a central role in selecting a (graded and ordered) set of relevant/salient properties that guides lexical access, which conforms to results of object naming experiments: "Assuming that there is a close relationship between gaze and visual attention […] we found that speakers have a strong tendency to fixate on each object they name and that the order of looking at the objects corresponded almost perfectly to the order of naming. Most importantly, we found that speakers fixated on each object until they had retrieved the phonological form of its name." (Levelt et al. 1999: 69).

For several reasons, therefore, it is (bottom-up) selective attention that *determines* the (order of the) aspects to be verbalized and their linguistic realization. The selection process may be modulated by top-down aspects of (pragmatic) knowledge, intention, attitude, context etc. (again compare the model of Itti & Koch 2001 for the case of visual attention). Overall, it realizes the subtle "perspective taking" or *perspectivation* of the speaker.

In general, producing a text involves various decisions of the speaker/conceptualizer, e.g., which reference objects to introduce, maintain, or relate, or which static or dynamic aspects of the evolving scenario to consider. *Changes* of state drive a description forward, and with them the alleged origo of the conceptualizer (or not), compare for example "On our journey, we went/came to a small town. There/here we found…". During language production, therefore, actual perception and stored (non-)linguistic knowledge interact in moving the attentional pointers on multiple levels. It is assumed here that attention-based perspective taking is reflected in semantic representations, relevant both for the semantics of spatial expressions in general and for directional deictics' semantics in particular (see Carstensen 2023 for locative deictics' semantics).

4.3 The role of selective attention in a two-level semantics

Typologically, linguistic categorization of spatial information is known to differ, especially with respect to path and manner information according to the distinction of 'satellite-framed' (path information as syntactic complement as in English and German), 'verb-framed' (path inform-

ation coded in the verb) and 'equipollently-framed' (both path and manner are specified, each as a verb)[16] languages (Slobin 2005). Further variability has been shown above for English and German combinations of spatial verbs, particles and adverbs. A mixture can be found in Russian *idti(tje) sjuda* ('come here', lit. "go hither"), which corresponds to Longgu's COME as *la mai* 'go hither' (Wilkins & Hill 1995). Base verb combinations can also be intralexical, e.g., in Russian *prixodit* 'come' (telic particle *pri-+xodit* 'go'). Also, there are different verbs of "going", depending on whether there is actual walking-type (*idti*) or driving-type (*jechat*) movement or habitual walking-type (*xodit*) or driving-type (*jesdit*) movement.

Aside from the fact that "come" and "go" verbs cannot be regarded as lexical universals, the assumption of complex conceptual universals (COME, GO, BRING, TAKE) is also unjustified, as it is unclear whether COME entails locomotion or whether it contrasts with a change-of-location version of GET (see the Yucatec data and the cave scenario). This is especially relevant for the meaning overlap of COME and HITHER in German, and the analysis of the corresponding semantic contributions of verb and adverb/particle. Depending on how far the RP conditions have to be generalized overall, it is also questionable whether COME is necessarily deictic: while in Shibe the ground for the "coming" verb must always be the speaker location at utterance time (Nakazawa 2007: 78), there is no objective licensing condition in the German Sparta and "Hänsel and Gretel" examples.

Also, apparent cross-linguistic contradictions are in need of explanation. For example, Nakazawa reports that the translation of (23) requires a "go" verb in Chinese and a "come" verb in Japanese, respectively (while both are eligible in German to express different perspectives). "Bringing out" the garbage typically means 'AWAY from RP' in German, 'TOWARD the RP' in English. Finally, Wilkins & Hill (1995) found a discrepancy in the verbalization scenarios: Their scene 6 (bounded movement toward, but not reaching, speaker) is a perfect "come" verb instance in Arrernte, but can only be described with a (generic) "go" verb in Longgu.

(23) *Mr. Suzuki persuaded me to [go/come] to the boring meeting he chairs.*

It should therefore not come as a surprise that "[v]erbs of 'coming' and 'going' are not universal" (Levinson 2006: 117). In a cognitivist semantics, a semantic specification is regarded as a language-specific grammatical coding of some conceptual content, which by definition allows for cross-linguistic differences (e.g., English *wide* is coextensional with the German adjectives *weit* and *breit*). This many-to-many mapping between a semantic and a conceptual level (hence "two-level semantics", see Lang and Maienborn 2011, Carstensen 2023) already explains some of

[16] With deixis, languages like Thai even have three verbs within one clause (Zlatev 2007). Compare, however, the two-verb constructions in English *comes running* or German *kommt gelaufen* "comes run".

the variance. In a cognitivist *attentional* semantics as presented here, further differences (especially those involved in apparent intra- or cross-linguistic contradictions) may be identified as different PERSPECTIVATIONS, induced by selective attention.

5 Topics in the semantics of directional deictics

In the Fillmorean tradition, “come” and, in part, “go” verbs are regarded as expressing movement of a theme relative to some location that is characterized by some condition *generalizing* the conception of the speaker’s position as deictic center but still referring to some contextual elements (i.e., are *deictically anchored*) (see (8)). According to the collected evidence, however, movement is not a necessary condition (Yucatec, cave case) and the goal may be just a location changed to (*come* vs. *get*), the deictic condition may have to be generalized to a carte blanche (Hänsel and Gretel case), and both the source of the asymmetry (e.g., the influence of empathy) and the corresponding explanation for non-literal asymmetries (Clark 1974) remain unclear. Before presenting proposals for the semantics of directional deictics, I want to discuss some of the topics involved.

5.1 Abstractness

Especially in formal semantics, one often finds proposals for the semantics of spatial expressions that somehow refer to (simplified) aspects of the physical world (see Carstensen 2015). For example, local PPs are typically supposed to denote certain spatial regions, and path expressions are supposed to denote (temporally) indexed sequences of such regions. This is neither sufficient nor necessary, however. As to local PPs like *in the vase*, further represented knowledge of naïve physics is required to guarantee that fissures can only be located in the material parts of the vase, and water only in the hollow part, but not vice versa. For descriptions like *knot in the shoelace*, these proposals beg the question what the region should be in that case. With respect to motion, apparent motion examples show non-necessity and Yucatec change-of-location verbs show non-sufficiency of actual physical conditions for semantics. As Jackendoff writes, “[i]ncidentally, trajectories such as *to Boston* are often thought to intrinsically involve motion. A more careful analysis suggests that this is not the case” (Jackendoff 2019: 98). In general, therefore, spatial expressions are based on more abstract representations different from naïve representations of physical space.

Abstract uses of directional expressions can also be observed. “go” verbs may only involve dynamicity (e.g., clocks in German either “gehen” ‘go’ or “(bleiben)stehen” ‘stand(still)’) and/or linearity (of NON-ACTUAL LOCOMOTION, as in *The mountain range goes from Canada to Mexico*, see Carstensen 2019). “Come” verbs may only involve changes to some salient state (cp. *Winter*

is coming, or Lakoff's *Here comes the beep*). Such abstract uses are usually treated as metaphorical, which, however, shifts the problem to the mechanism of metaphorization. The alternative is to assume semantic representations with more abstract criteria in a two-level semantics.

The insight that semantics is only indirectly related to the world is the central point of Cognitivist attentional semantics (Carstensen 2011). Here, it is assumed that semantics primarily relates to representations of selective-attentional engagement in some cognitive domain and at a certain level (corresponding to the abstract representations mentioned). This has some *general* implications. First, it reflects the observation that language also codes *perspectivations* of the world rather than only unequivocal aspects of the world (cp. a "half empty/full" glass, or being a "few centimeters away from"/ "close to" something), as a representation of internal processing of external information. Second, it allows for subtle influences of language on cognition via culturally-determined linguistic preference for some attentional perspectivation (see the "pragmatic experience" above). Third, it offers an explanation for metaphors/metaphorical shifts of meaning via common attentional structure across domains. Fourth, it allows to treat the semantics of deictic expressions via their content-leanness and by sole reference to attentional aspects (see Carstensen 2023), as opposed to analyses based on concepts like 'deictic center', 'pointing', 'motion' etc.

5.2 Denotation ranges and lexico-syntactic competition

It was Saussure who already maintained that the "semantic value" (the denotation) of a term depends in part on its lexical siblings in some language. Such influence of the lexico-syntactic system as a whole on the meaning and use of individual lexemes – famously shown for cross-linguistic variation of color terms – can perhaps also be found in directional deixis.

In the case of causative directional deictics (where German and English senses differ), the German comitative verbs *mitbringen* 'Theme moves (with) object to RP' and *mitnehmen* 'Theme moves (with) object away from RP' represent the characteristic senses of deictic, centripetal *bring*, and centrifugal *take*, respectively. Deictic direction can also be expressed with *herbringen* 'bring hither' or *hinbringen* 'take thither'. This could be the reason why German *bringen* itself is directionally neutral/underspecified (contra English *bring*, see Preston 1984, but cf. Hockett 1990). In non-deictic use, it is typically rather centrifugal, whereas the centripetal sense of *bring* is usually expressed by *holen* 'get/fetch' (most evident by the meme-gone request "Hol mir mal ne Flasche Bier" 'Get/bring me a bottle of beer' by German ex-chancellor Schröder).

In Longgu, there is no lexical contrast of "coming" and "going", but only the *la* 'go' verb – covering both directions – with its adverbial spatial deictic contrast. As Wilkins & Hill (1995) show, however, this has repercussions for the denotational ranges of "come" expressions: There are

twice as many unambiguous COME scenes in Arrernte (where movement *toward* the deictic center is sufficient) as in Longgu (where movement is expected to end only at the deictic center), and the range of unambiguous “go” scenes is also wider in Arrernte.

Also, the (non-)presence of a lexeme carrying some aspect of expressive/social meaning or requiring some context of use (register) may have an impact. For example, German generally lacks a neutral/casual change-of-location “get” verb (but rather has higher-register *gelangen* and negative *geraten*, see above). This may be the reason for the (wider) use of *kommen* in the cave scenario.

5.3 Semantic composition

Above (see Figure 2 and Figure 3), it was proposed that spatial and eventuality *aspectual structures* of some situation must, at least in satellite-framed languages, be compatible. English *come here*, like its Russian translation, seems to rebut such an isomorphism. However, usage cases like these may have to be analyzed as expressions in which the adverbial is simply not marked as directional (compare *come *(to) me, come in(to) the garden*).[17] For satellite-framed languages, this would support the hypothesis that verbs basically code eventuality-level information, whereas spatial-level information is basically coded in PPs, adverbs, and particles.

Bohnemeyer’s proposal for a change-of-state (location) semantics in Yucatec essentially entails a non-representation of spatial motion and continuity (which we know from apparent motion is cognitively given) on the eventuality level. Again, this is not necessarily exotic: change of location as abstraction from the continuity of a figure’s path can be found in causative position verbs (compare English *put the vase on the table*) or in specific situations (cave-scenario). This, however, means that there is no continuity (i.e., no path semantics of spatial directional complements) on the spatial level either (see Carstensen 2019).

5.4 Deictic reference

Wilkins & Hill already pointed out that “go” verbs are not inherently deictic: “if there is a deictic opposition manifested by COME and GO forms, it is at the level of pragmatic interpretation” (Wilkins & Hill 1995: 251). While they argue with their field studies’ results, lexical aspect is restricted to achievement/accomplishment for *come/kommen* in English/German, but unrestricted (or underspecified, see Lukassek et al. 2017) for *go/gehen* (compare the unbounded activity

[17] A similar case is the abstract use of German *das kommt daher* ‘this is because…’, lit. “that comes there-hither”. It has to be analyzed as *das kommt VON da her* “that comes FROM there hither”. For the same reason, spatial *dahin (thither)* means ‘TO there thither’, *woher (whence)* ‘FROM where hither’, and *wohin (whither)* ‘TO where thither’.

**Come/Go for an hour through the woods*). Directional use of *go* can therefore be non-deictic (*Go to the store (from wherever you are).*, *Go!*), and is (only) source-oriented when deictic (*Go (away from me)!*).

Accordingly, *go* is a dynamic verb that can have different aspectual characteristics depending on complementation or pragmatic use. *Come*, on the other hand, is bounded, goal-oriented, and underspecified for the pre-change condition. Hence it contrasts not only pragmatically with *go*, but also semantically with *get*. Additionally, given the reported evidence, some languages (e.g., Shibe) may require specific deictic reference lexicalized as a semantic condition for their "come" verb, whereas others only seem to require a change to a result state (deictic reference being a characteristic/prototypical condition).

5.5 Deictic framing

In discussions of directional expressions, the distinction of eventuality-level and spatial-level information is often neglected in favor of the typological framing distinctions. For the framing of deixis, which is often presented as a single component, this presents a problem. Compare the child's verbalized intention to join dinner in English *Coming (*hither)!*. It shows that RP-at-goal *come* and RP-at-goal *hither* do not necessarily match. Similarly, in the question (24) (after having opened the cave door), the alleged RP-at-goal "come" verb is perfect in German but only compatible with a RP-at-source adverb *wohin*.[18]

(24) *Wohin/*Woher* *kommt* *man* *da?*
(to) where-thither/*-hither comes one there?
'Whither/where does one get there?' [looking into the cave]

These observations can be explained by assuming that verb and adverb do not straightforwardly relate to the (same) deictic center, but frame (different) perspectivation on different levels, with eventuality-level information being less granular and/or specific than spatial-level information. Such an analysis is corroborated by the German three-component specifications, which are not redundant in (19) and (20), and not necessarily conflicting in (21) and (24). While directional deictic verbs are RP-related (if at all), directional deictic adverbs typically are *speaker*-related (compare *Coming (*hither)!*). In general, however, the semantics of spatial deictics is much more abstract and also allows non-RP-related uses. For the spatial adverbs, this is the case with German *hin und her* "to and fro/hither and thither" '(iteratively) in some direction (possibly transverse) and back'.

[18] See Footnote 17 on why *woher* is unacceptable here.

5.6 Valuation

Based on the previous discussion, it can be argued that the research question underlying the semantics of "coming" verbs should be rephrased from "What are the spatial goal conditions for *come* (as opposed to *go*)?" to "What are the conditions for a speaker to select *come* as a CHANGE verb with a distinguished (spatial) result state?". EMPATHY with a speech act participant seems to be relevant for this in some languages (Kuno & Kaburaki 1977).

It has been observed, however, that *structural* aspects also play a role here in that the empathy for the subject of some matrix verb may exceed the empathy for the speaker, who has the highest empathy rank by default (compare the contrast in (23)). "Thus in Japanese, the lexical meaning of *kuru* 'come' is amended with an additional clause: 'MOVE TOWARD a point which is the location of the referent of the matrix subject when embedded under control predicates'" (Nakazawa 2007: 73). And yet, such expressions may simply reflect a culturally determined structural impact on the PROMINENCE of (and empathy with) some matrix verb's subject. According to such a principle-based description, the conditions for the use of "coming" verbs would simply involve a relatively higher prominence-based valuation of the result state, influenced by different factors and giving room for contextual and cross-linguistic variation.[19]

Different from other approaches that use the notion 'prominence' for perspectival expressions (Anderson 2021) as an objective description of an element's salience, I will use RELEVANCE to emphasize the subjectivity of the valuation for the speaker in the current context. "Relevance" thus refers to the sub-intentional, cumulative subjective salience of elements (which is akin to the empathy proposals, but distinct from the QUALITATIVE valuation of the norm- or evaluation-based analysis of abstract directionals by Clark). Under the assumption that salient states attract attention, VALUATION of result states based on objective or subjective criteria is proposed here as a central, general factor determining "coming" expressions. Such RELEVANCE-BASED valuation might even explain the use cases of *come* not yet accounted for (*come unraveled* etc.).

Fillmore (1975: 67) gives an example for the context-dependence of valuation. He says that one can talk about an uninhabited island in a lake, and about a loon "coming" there at night (with RP backgrounded), but that this cannot be followed by a sentence like *I would like to come there some day* (explained here with higher valuation of the source-state due to the involvement of the speaker).

[19] Compare the differential use of *come* and *go* in the study systematically varying some factors by Groussier reported in Fortis & Fagard (2010: 24). Fillmore also mentions that in Mexican Mazahua, "one uses the place deictic words with the poles reversed" (Fillmore 1975: 83) in deferential language, which corresponds to an upvaluation of the addressee.

6 The semantics of directional deictics

6.1 General aspects

As has been shown, the semantics of directional deictics cannot be based on movement and generalizations of the deictic center. These concepts are rather only relevant in the prototypical spatial cases of deictics. Instead, it must be based on a simple, abstract change and a more abstract characterization of the deictic oppositions. The basic idea is to acknowledge the relevance of internal, attentional processing of (and even independent of) the representation of the (projected) external world. The general aspects of the proposal made here are the following.

First, the classic pair of contrasting “come”/“go” verbs has to be augmented with a third, “get”, verb. While “come” is underspecified with respect to an activity (which it may share with “go”) in its preparatory phase, and therefore has accomplishment and achievement interpretations, it directly contrasts with “get”, but only indirectly/pragmatically with “go” (which expresses an activity, but is underspecified for spatial aspect).

Second, directional deictics are assumed to express a change of states, generalizing over spatial and non-spatial meanings. The dimension on which the spatial changes are based is derived from attentional scanning (i.e., successive attentional engagement on some representational content), either exogenously driven (i.e., by tracking movement, both in objective and subjective time) or endogenously driven (by tracing linear entities in non-actual locomotion cases, exclusively in subjective time), see Carstensen (2019). Clark’s abstract cases involve changes of state based on the temporal development of an object’s property over time (*He went mad*).

Third, instead of generalizing the criteria for “come” or “hither”, a different, abstract condition seems to be required. It is proposed here that deictic contrast is realized as represented attentional focus (of the conceptualizer) on one of the states of the pertinent change, which is coined as a semantic condition.

Fourth, attentional selection of one of two options therefore corresponds to perspective-taking, determined by the valuation of a state´s content for the speaker. Here we have Fillmore’s criteria, but only as conceptual conditions involved in computing the aggregate salience value of a state. This attentional perspectivation reflects the speaker’s origo (and its alleged shifts) in the required generalized sense, i.e., the (spatial) state the speaker associates with most at the moment of conceptualization, but only in prototypical cases with her actual position at the goal/source for “come”/“get/go”. Non-literal use of deictic verbs may involve additional objective (norm(ality)-related) or subjective (evaluative) valuation in some languages.

Fifth, directional verbs and adverbs are seemingly alike in their semantic contribution of path-related information (and their compatibility must be ensured in semantic composition), but they

may still express different perspectives (see the German data (19)-(21)). This is explained here with attentional perspectivation on different (the eventuality and spatial) levels, and with the incremental process of language production.

6.2 Eventuality-level directional semantics

I will now discuss the *specific* implications of the previous considerations (and with selected expressions as a model for others), starting with the proposal for the semantic entry of English *go* in (25). It shows a three-place predicate, where p is the spatial-level (path) argument, x corresponds to the theme/figure, and e is the eventuality-level referential argument. The eventuality e is characterized by three conditions: “DYNAMIC(x)” represents a necessary endogenous or exogenous micro-change, to distinguish *go* from static verbs like *stand*. “SCANNED(x)” represents sequential attentional scanning of the figure, which is either tracked as an actually moving object or traced as in non-actual locomotion. “MOVEMOD(x)” is a generalized mode-of-movement predicate (instantiated in *walk*, *drive* or specified with *by car* etc.) that indicates actual spatial locomotion usage of *go* of some kind. The LINK predicate signifies correspondence of eventuality-level and spatial-level aspectual structure in this satellite-framed semantics (see the discussion in Carstensen 2019: 39 on that point).

(25) ⟦ *to go* ⟧:
λp λx λe [[DYNAMIC(x) & SCANNED(x) & MOVEMOD(x)](e) & LINK(e,p)][20]

The proposed semantic entry for *come* in (26) is strikingly different. It is marked as a change verb with a post-change state of attending to the figure/theme (“ATT(x)”), whose pre-change eventuality-type is underspecified (hence is compatible with movement or change of location). “VALUED” signifies the role of qualitative valuation for this verb (but would be missing for German *kommen*). Evidently, attention figures on two levels, both defining a state of the figure being attended and the goal state as being attended. The latter marks the difference to the contrasting term *get* and its semantic entry in (27), where the focus is on the pre-change state. Anchoring is realized here as presuppositional content (signified by underlining) involving the currently relevance-valued focused state, either the result/goal state (*come*), or the initial/source state (*get*).

(26) ⟦ *to come* ⟧:
λp λx λe [[[ATT(x)](s2) & ATT(s2) & VALUED(s2) & CHANGE(_, s2)](e) & LINK(e,p)]

(27) ⟦ *to get*$_{\leftrightarrow come}$ ⟧:
λp λx λe [[[ATT(x)](s1) & ATT(s1) & CHANGE(s1, _)](e) & LINK(e,p)]

[20] To distinguish the levels, further conditions may have to be added (“EVENTUALITY(e) & SPATIAL(p)”). Note also that general aspects/problems of semantic structure are not discussed here.

In other words, it is assumed here that *come* and *get* contrast semantically in their focus on different states of an eventuality-level change, whereas *come* and *go* contrast indirectly/pragmatically: if the semantic conditions for *come* (and *get*) are not satisfied, *go* expresses the mere activity of "getting away" from the source in a deictic context (i.e., RP at source). Note that eventuality-level attention in "come"/"get" semantics lastly identifies the lexical differences as one of (attentional) perspectivation, that is, as a subjective criterion that only reflects objective aspects (if at all, see (15b) and (16b)).

Non-objectivity is apparently what also seems to matter in the "horrifying" examples of Fillmore repeated in (28): although both sentences describe the same situation, only (28b) is rejected by a "large number" (Fillmore 1975: 68). The same holds for the data presented by Kuno & Kaburaki (1977), especially their (29). While Fillmore is startled by the presence of multiple deictic centers in a sentence, Kuno & Kaburaki complain about the corresponding violation of their principle ("Ban of Conflicting Empathy Foci"). The problem dissolves if the sentences are seen as subjective construals with dynamic perspectivations in language production. Such dynamic attentional construals beyond objective conditions would involve a valuation shift from the addressee's home to the higher-ranked speaker's home (both licensing *come*) in (28a), whereas relevance-based focus will "stick" to the speaker's home base in (28b) (to be rather verbalized as "after he gets/goes to your house").

(28) [different houses are ground, from Fillmore (1975: 68)]
   a. *He'll come to your house before he comes to my house.*
   b. *??He'll come to my house a*

(29) [same party is ground, from Kuno & Kaburaki (1977: 663)]
   a. *I will not go to a party that John is going to.*
   b. *I will not go to a party that John is coming to.*
   c. **I will not come to a party that John is going to.*
   d. *I will not come to a party that John is coming to.*

Similarly, in (29b), once the party gets relevance-based focus, John can be said to be "coming" to it. Yet focusing on the party location is a consequence of *subjective* conceptualization, as shown by (29a) (in that case, the actual/utterance location of the speaker is higher-ranked). Suffice it to say that this proposal directly corresponds to the hypothesis of Kuno & Kaburaki (1977) that (29b) "is acceptable because it involves a natural psychological shift from the place of utterance to the party" (p. 663). As the contrast of (29c) and (29d) shows, there is nothing to shift if an element already has the highest relevance value.

6.3 Spatial-level directional semantics

The proposed semantics for the German directional adverbs in (30) and (31) are structured analogously to those of *come* and *get* as a change of attentional states, only on the spatial level.

(30) ⟦ *her* ⟧: λp [[[ATT(x)](s2) & ATT(s2) & CHANGE(_, s2)](p) & SPATIAL(p)]

(31) ⟦ *hin* ⟧: λp [[[ATT(x)](s1) & ATT(s1) & CHANGE(s1, _)](p) & SPATIAL(p)]

It is one of the central – and correspondingly foregrounded – observations of Wilkins & Hill (1995) that a “come” expression (*petye-*, containing the morpheme *-tye-* ‘hither’) is used in Arrernte where a “go” expression would appear in other languages (a figure only getting closer to the deictic center but not reaching it). Given the data presented here, this is less astonishing if in Arrernte, the moving figure’s approach at some point is regarded as the relevant factor for subjective eventuality-level valuation. Longgu is clearly different. As shown in the Wilkins & Hill paper, its “come” expression (*la mai* ‘go hither’) relates to the deictic center at the goal of movement. This is explained here by the fact that change and valuation/focus happen at the spatial level reflected by the adverb (i.e., without more abstract eventuality-level conceptual criteria).

The German three-component data can be interpreted along this line of argumentation. *Kommen* involves relevance-based valuation of the goal state and characterizes a roughly specified goal area. *Her*, corresponding to Longgu *mai*, is much more restricted (typically to the actual speaker-centered area). This is the reason why *kommen* is also compatible with *hin* (*Komm hier/da hin!*): *hin* signifies an attentional source-focused dislocation to a finer-granular location on the spatial level. Within this location, primary/unspecific/secondary attention to the place of the figure can be differentiated with *hier*/*da*/*dort* ‘here/there/over there’ (see Carstensen 2023). It is important to note that different perspectivations on the levels are possible for *kommen* and *hin*, but also for *gelangen* ‘get’ and *her*: *Wie seid ihr hierher gelangt?* “How did you get hither/here?” is perfect.

This analysis holds a processing-oriented view according to which verbalization may start with verb production, followed by modifiers and arguments (as has been demonstrated for quantification in Carstensen 2021). The production perspective also helps to explain the differing judgments of combinations in (21). *?Geh hier her/hin!* is strange, because monitoring and feedback loops in generation (see Levelt 1989) would lead to the detection and repair of inadequate expressions under normal conditions. In this case, if there is movement to the speaker, *come* should be selected as the more specific term (see the hyperonym problem above). **Komm da her!* is even worse, since *da* is in conflict with *komm+her*, evidenced by the fact that *Komm her!* pragmatically implies KOMM HIER HER!. Finally, **Komm hin!* is unacceptable because it leaves

the information about the exact location of the figure unspecified (as it is different from *Komm her!*, and thus excludes HIER pragmatically).

7 Conclusion

Classical approaches to the semantics of directional deictics use specific domain (i.e., spatial) and content (e.g., deictic center, location of speaker, home base) criteria, or generalizations thereof, to characterize the meaning of directional deictics. Not only have they overlooked the change-of-location semantics of the “come” and “get” verbs and the corresponding inadequacy of a simple, movement-based “come”/“go” contrast, they also fail to capture the abstract, subjective or valuative senses of directionals, the subtle differences in cross-linguistic comparisons, and the astonishing variation of acceptability ratings in complex sentences.

The attentional approach to spatial semantics as presented here assigns a primary role to represented operations of selective attention. With attentional layers distinct from domain and content, it offers the abstract criteria needed to more generally capture the various uses of directional deictics (as has been proposed for the semantics of locational deictics). Selective attention has two functions in directional semantics: It *defines* eventualities as being static or dynamic via constant (ATT) or permanently changing (SCANNED) attention to elements, respectively, and it *characterizes* one of the states of a change as the one being in focus. In a two-level scheme of semantic interpretation (motivated, e.g., in Carstensen 2023 and Wilkins & Hill 1995), abstract language-specific semantic representations are instantiated by some conceptual content in a given context, resulting in a specific utterance meaning of the terms. From a general Cognitivist viewpoint, furthermore, semantic and conceptual representations are bidirectionally connected, and their relationship is shaped by linguistic usage in a language (the “pragmatic experience” of Gerwien & Stutterheim 2022b).

It turns out that, for English, *come* and *get* are the proper contrasting terms on the eventuality level, as they express a change to or from a state, respectively. Depending on the dynamicity of the other state, they refer to simple changes of locations or to extended movements (*come here, get away*). In these respects, *come* and *get* correspond to German *her* and *hin* on the spatial level, only that the latter are defined by the dislocation of the attentional focus in the spatial domain alone. Also, and in line with Wilkins & Hill (1995), *come* is seen here as only indirectly contrasting with *go* – at least in this satellite-framed language. Rather, *go* is a dynamic verb that is selected if the *come* and *get* criteria are not satisfied. In prototypical cases, it refers to an extended, potentially laborious action, while *come* refers to changing to a relevant goal state. This may be the reason why *go* and *come* are used to non-literally/metaphorically express (negative) development and (positive) outcome, respectively.

Typical typological presentations of directional deictics suggest that DIRECTIONAL DEIXIS is a single conceptual aspect whose differential appearance in language only depends on framing (i.e., how it is incorporated into linguistic meaning). Instead, it is proposed here that the relevance of the goal state may be interpreted differently on the eventuality (conceptual) and spatial level (*Coming (*hither)!*), which explains the German multi-component data and the different goal(-state) scopes in cross-linguistic comparisons.

Perspectivation differences in deictic contrasts on some level can therefore be traced back to which state of a corresponding change is in focus at the time of linguistic categorization (represented as presuppositional content in the semantic specification), influenced by relevance-based valuation of the states (based, in turn, on various aspects, e.g., Fillmore's criteria). In that respect, the attentional approach both encompasses and surpasses previous explanations of concrete, abstract, and non-spatial uses of directional deictics, and provides an alternative to using too specific or rigid criteria (deictic center, movement, empathy hierarchy etc.). In particular, valuation-based perspectivation is subjective and context-dependent, solving some of the problems Fillmore and Kuno & Kaburaki have with their data.

There is at least one riddle left, however. I have reported Rauh's insight that German *kommen* 'come' lacks a semantic condition of subjective or objective qualitative valuation, which makes the verb applicable to a larger set of result-state situations with goal perspectivation (Rauh 1981). Fortis and Fagard mention *Er kam vors Kriegsgericht* "He came before the war council", involving an undeniably negative goal state (Fortis & Fagard 2010). This generality could therefore account for *kommen*'s being acceptable in the cave scenario, corresponding to the peculiarity of German verb *bringen* 'bring/take' in being underspecified with respect to deixis and attentional perspectivation. The question is: could *kommen* – unlike goal-focused *come* – also be a regular expression for a *source*-focused perspectivation of a *change of location*, as a filler of a lexical gap (i.e., as a casual term alongside *gelangen* and *geraten*)? In that case, *kommen* 'come' would sometimes mean 'get'.

Acknowledgements: The author wishes to thank two anonymous reviewers for their positive feedback.

References

Anderson, Carolyn Jane. 2021. Coming in, or going out? Measuring the effect of discourse factors on perspective prominence. *Experiments in Linguistic Meaning* 1. 1–14. https://doi.org/10.3765/elm.1.4865.

Barlew, Jefferson. 2017. The semantics and pragmatics of perspectival expressions in English and Bulu: The case of deictic motion verbs. Ph.D. Thesis. The Ohio State University.

Bohnemeyer, Jürgen. 2010. The Language-Specificity of Conceptual Structure: Path, Fictive Motion, and Time Relations. In Barbara Malt & Phillip Wolff (eds.), *Words and the Mind: How words capture human experience*, 111–137. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195311129.003.0007.

Bühler, Karl. 1990[1934]. *Theory of language. The representational function of language*. Amsterdam & Philadelphia: Benjamins.

Carstensen, Kai-Uwe. 2011. Toward cognitivist ontologies. *Cognitive Processing* 12(4). 379–393.

Carstensen, Kai-Uwe. 2013. A cognitivist semantics of gradation. *Zeitschrift für Sprachwissenschaft* 32(2). 181–219. https://doi.org/10.1007/s10339-011-0405-0.

Carstensen, Kai-Uwe. 2015. A cognitivist attentional semantics of locative prepositions. In Giorgio Marchetti, Giulio Benedetti & Ahlam Alharbi (eds.), *Attention and meaning. The attentional basis of meaning*, 93–132. Hauppauge, NY: Nova Science Publishers.

Carstensen, Kai-Uwe. 2019. From motion perception to Bob Dylan. A cognitivist attentional semantics of directionals. *Linguistik Online* 95(2). 17–50. https://doi.org/10.13092/lo.95.5514.

Carstensen, Kai-Uwe. 2021. Quantification: The view from natural language generation. *Frontiers in Artificial Intelligence* 4. https://doi.org/10.3389/frai.2021.627177.

Carstensen, Kai-Uwe. 2023. Attentional Semantics of Deictic Locatives. *Lingua* 296. https://doi.org/10.1016/j.lingua.2023.103626.

Clark, Eve V. 1974. Normal States and Evaluative Viewpoints. *Language* 50(2). 316–332. https://doi.org/10.2307/412440.

Fillmore, Charles J. 1975. *Santa Cruz Lectures on Deixis 1971*. Bloomington, IN: Indiana University Linguistics Club.

Flecken, Monique, Panos Athanasopoulos, Jan Rouke Kuipers & Guillaume Thierry. 2015. On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition* 141. 41–51. https://doi.org/10.1016/j.cognition.2015.04.006.

Fortis, Jean-Michel & Benjamin Fagard. 2010. Space and language. Course Presented at the Leipzig Summer School on Linguistic Typology. Part V – Deixis.

Gerwien, Johannes & Christiane von Stutterheim. 2022a. Conceptual blending across ontological domains – References to Time and Space in motion events by Tunisian Arabic speakers of L2 German. *Frontiers in Communication: Language Science* 7. 856805. https://doi.org/10.3389/fcomm.2022.856805.

Gerwien, Johannes & Christiane von Stutterheim. 2022b. Describing motion events. In Andreas H. Jucker & Heiko Hausendorf (eds.), *Pragmatics of Space*, 153–179. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110693713-006.

Goddard, Cliff. 1997. The semantics of coming and going. *Pragmatics* 7(2). 147–162.

Graziano, Michael S. A. 2020. Consciousness and the attention schema: Why it has to be right. *Cognitive Neuropsychology* 37(3–4). 224–233. https://doi.org/10.1080/02643294.2020.1761782.

Hockett, C. F. 1990. Bring, take, come, and go. *Journal of English Linguistics* 23(1–2), 239–244. https://doi.org/10.1177/0075424290023001-219.

Itti, Laurent & Christof Koch. 2001. Computational modelling of visual attention. *Nat Rev Neurosci* 2(3). 194–203. https://doi.org/10.1038/35058500.

Jackendoff, Ray. 2019. Conceptual semantics. In Claudia Maienborn, Klaus Heusinger & Paul Portner (eds.), *Semantics – Theories*, 86–113. Berlin & Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110589245-004.

James, William. 1890. *The principles of psychology*. New York: Henry Holt and Company.

Kohler, Axel, Leila Haddad, Wolf Singer & Lars Muckli. 2008. Deciding what to see: The role of intention and attention in the perception of apparent motion. *Vision Research* 48(8). 1096–1106. https://doi.org/10.1016/j.visres.2007.11.020.

Kuno, Susumu & Etsuko Kaburaki. 1977. Empathy and syntax. *Linguistic Inquiry* 8 (4). 627–672. http://www.jstor.org/stable/4178011.

Lang, Ewald & Claudia Maienborn. 2011. Two-level semantics: Semantic form and conceptual structure. In Klaus von Heusinger, Claudia Maienborn, & Paul Portner (eds.), *Semantics. An international handbook of natural language meaning* (Vol. 1, HSK 33.1), 709–740. Berlin & New York: De Gruyter.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge: MIT Press.

Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–75. https://doi.org/10.1017/S0140525X99001776.

Levinson, Stephen C. 2006. Deixis. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics* (Blackwell Handbooks in Linguistics 16), 97–121. Malden, MA: Blackwell publishers.

Lukassek, Julia, Anna Prysłopska, Robin Hörnig & Claudia Maienborn. 2017. The semantic processing of motion verbs: Coercion or underspecification? *Journal of Psycholinguistic Research* 46. 805–825. https://doi.org/10.1007/s10936-016-9466-7.

Malsch, Derry L. & Kathleen M. Lant. 1977. On 'Normal State' deixis. *Linguistic Inquiry* 8(4). 744–746.

Mitchell, Margaret, Kees van Deemter & Ehud Reiter. 2013. Generating expressions that refer to visible objects. *Proceedings of NAACL-HLT 2013*. 1174–1184.

Nakazawa, Tsuneko. 2007. A typology of the ground of deictic motion verbs as path-conflating verbs: The speaker, the addressee, and beyond. *Poznań Studies in Contemporary Linguistics* 43(2). 59–82. https://doi.org/10.2478/v10010-007-0014-3.

Oshima, David Y. 2006a. GO and COME revisited: What serves as a reference point? In *Proceedings of the 32nd annual meeting of Berkeley Linguistics Society*, 287–298. Berkeley: University of California.

Oshima, David Y. 2006b. Perspectives in reported discourse. Ph.D. thesis. Stanford University.

Preston, Dennis R. 1984. Take and bring. *Word* 35(2). 177–186. https://doi.org/10.1080/00437956.1984.11435755.

Ramachandran, Vilayanur S. & Stuart M. Anstis. 1986. The perception of apparent motion. *Scientific American* 254(6). 102–109. https://doi.org/10.1038/scientificamerican0686-102.

Rauh, Gisa. 1981. On coming and going in English and German. *Papers and Studies in Contrastive Linguistics* 13. 53–68.

Simons, Daniel J. & Christopher F. Chabris. 1999. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception* 28(9). 1059–1074. https://doi.org/10.1068/p281059.

Slobin, Dan I. 2005. Linguistic representations of motion events: What is signifier and what is signified? In Costantino Maeder, Olga Fischer & William J. Herlofsky (eds.), *Outside-in – inside out* [Iconicity in Language and Literature 4], 307–322. Amsterdam/Philadelphia: John Benjamins.

Sudo, Yasutada. 2018. *Come vs. go and perspectival shift.* New Brunswick, NJ: Handout for talk at Rutgers University Department of Linguistics.

Tanenhaus, Michael K., Michael J. Spivey Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–1634. https://doi.org/10.1126/science.7777863.

Wälchli, Bernhard & Fernando Zúñiga. 2006. Source-Goal (in)difference and the typology of motion events in the clause. *Language Typology and Universals* 59(3). 284–303. https://doi.org/10.1524/stuf.2006.59.3.284.

Wilkins, David P. & Deborah Hill. 1995. When “go” means “come”: Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6(2/3). 209–259. https://doi.org/10.1515/cogl.1995.6.2-3.209.

Zlatev, Jordan. 2007. Spatial Semantics. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 318–350. Oxford: OUP.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 6
https://doi.org/10.18148/zs/2026-2006

Roland Hinterhölzl, University of Venice, rolandh@unive.it

Privileged Access, V2 and the *that-trace* effect

Abstract: The paper discusses the famous that-trace effect and the strategies used to circumvent it in German, English, French and Italian. While the data have been widely discussed in the literature, the paper proposes a new account of the phenomenon underlies these strategies and argues that a violation of economy in subject movement is at the core of the that-trace effect. In particular, it is argued that the that-trace effect targets subjects and not objects or adjuncts, since subjects stand out in having privileged access to the C-domain. Cases in which subjects to not give rise to that-trace effects and cases where other elements than subject do give rise to a that-trace effect, will be explained as falling out from grammatical system of anchoring arguments to the context.

Keywords: that-trace effect, high and low V2, extension of the phase edge, economy of derivation, referential anchoring, big DP-hypothesis

Cite as: Hinterhölzl, Roland. 2026. Privileged Access, V2 and the *that-trace* effect. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi.org/10.18148/zs/2026-2006
Received: 03 August 2025; **accepted:** 02 June 2026; **published:** 30 June 2026
Published by: publikon / dgfs (Linguistic Society of Germany)

1 Introduction

This paper presents a novel approach to the *that-trace* effect that is based on the insight that subjects stand out in having privileged access to the C-domain. Paradoxically, this privilege of subjects – that is visible in various effects appearing with local, that is, clause-internal movement – leads to a handicap in long distance movement. I argue that the effect is due to an economy condition operative with flexible phase edges in the C-domain. Flexible phase edges are also visible in V2-languages, which distinguish each other between low and high V2, corresponding to a low (FinP) or high (ForceP) phase edge in the C-domain.

The paper, moreover, is inspired by Pesetsky's (2017) observation that the *that-trace* effect shows itself in various different effects in a number of (related and unrelated) languages. The present paper focuses on the strategies circumventing the *that-trace* effect in the four languages that I know best: English, German, Italian and French. It will thus not provide new data. Almost all the empirical phenomena dealt with in the paper have been already widely discussed in the literature. What is novel or innovative in the paper, is the fact that a number of well-known phenomena are bound together under a common approach that throws a new light on the nature of *that-trace* effect.

It is largely unexplained why subject extraction in English should give rise to the *that-trace* effect, as is illustrated in (1ab). First, it is unclear why adjunct extraction should be 'easier' than extraction of an argument and second, it remains mysterious (at least in my opinion) why deletion of the complementizer circumvents ungrammaticality. Nevertheless, the *that-trace* effect is a robust phenomenon that appears in many languages, including Russian and Standard Arabic, as is pointed out by Pesetsky (2017). Moreover, Pesetsky (2017) shows that it is connected with a wide range of phenomena that allow the circumvention of *that*-deletion, including the French *que*/*qui*-rule, complementizer agreement in West Germanic, and (evidence for) low subjects in Italian, to name the cases most discussed in the literature. Even in English, the *that-trace* effect can be circumvented, as is also pointed out by Pesetsky (2017), if an adjunct intervenes between the complementizer and the trace, as is illustrated by the contrast between (1a) and (1c), having given rise to prosodic accounts of the effect (cf. Kandibowicz 2006, 223).

(1) a. **Who do you think that t met Mary* (*that-trace* effect)
b. *Who do you think met Mary* (deletion of the complementizer)
c. *Who do you think that in this office t is the boss* (lack of the effect)

Within the literature regarding the Government and Binding theory, accounts focused on the position from which subjects are extracted, namely [Spec,TP], as a non-lexically governed position (cf. Chomsky & Lasnik 1977). There are two newer proposals – besides Sobin (2002), discussed in Section 6 below – that tackle the *that-trace* effect within the minimalist program.

First, there is the account by Bayer and Salzmann (2013) that argues that the *that-trace* effect is not limited to subjects but pertains to the highest argument in the clause. In particular, they propose that the *that-trace* effect results from improper movement of the aboutness topic in the embedded clause. I will argue below that the *that-trace* effect pertains to the highest element that serves for anchoring the (embedded clause) in the context. Second, there is the account by Padovan, Bidese and Tomasello (2021), who argue that Nominative case assignment or the lack of it is a major factor in explaining the absence of the *that-trace* effect in Cimbrian and Bavarian. However, their account is limited to these varieties and has nothing to say, for instance, about the *that-trace* effect in English.

I will argue below that the *that-trace* effect in essence follows from an economy condition connected with the bottleneck effect in [Spec,FinP] and discuss the strategies that are available in English, German, French and Italian to circumvent the *that-trace* effect, showing that all these reduce to extending the phase edge from [Spec,FinP] to [Spec,ForceP] in the embedded clause.

2 Privileged access and restrictions on movement

I will provide two examples to show that subjects have privileged access to the C-domain, one from English and one from German. As far as the structure of the C-domain is concerned, I will follow Rizzi (1997) in assuming that various discourse heads, licensing Topics and focused constituents, occupy the area between FinP, which I analyze as constituting the lower edge of the C-domain and ForceP, constituting the higher edge of the C-domain, as is illustrated in (2).[1] Furthermore, I assume that topics above Focus are base-generated, while the topics below Focus are derived by movement (cf. Hinterhölzl 2021 for arguments). A note on the nature of Force^0 is in order: while in embedded clauses it encodes sentence type (for instance, interrogative or non-interrogative), I take it to encode illocutionary force in main clauses. But this issue is not important for the arguments in the present paper.

(2) $[_{\text{higher edge}}$ Force Topic* Focus Topic* $[_{\text{lower edge}}$ Fin [TP ....]]]

2.1 Locality conditions on movement

I assume that movement across domains is restricted by phases, following a standard assumption in the minimalist program (cf. Chomsky 2000), as stated in (3). As is illustrated in (2), the C-domain has two edges, raising the question of whether both phase edges are relevant for regulating movement into and out of the C-domain.

[1] An anonymous reviewer asks whether the account is tied to an extended C-domain, as given in (2) or would also work in a simpler system with CP-recursion (cf. Vikner et al. 2017). As far as I can see, this would be in principle feasible, aligning the lower CP with low V2 and the higher CP with high V2.

(3) Locality: XP-movement is limited from one phase domain to the adjacent phase domain by touching down in the phase edge.

I will argue below that the C-domain has only one phase edge that – based on the famous bottleneck effect in V2-languages – is identified by default with its lower edge. However, I will also argue that this single phase edge is flexible. A flexible phase edge means that the phase edge in FinP can be extended to ForceP under a specific condition to be discussed below. Moreover, I will argue that phase extension is indicated (made visible) by movement of the exponent of the lower edge head to the head of the higher edge.

In other words, I am reviving the notion of phase extension in den Dikken (2007), which he used to account for the derivation of predicate inversion and for dative alternation, that is, for movement out of the v-domain. In his work, phase extension from an inherent phase – defined in terms of predication – to a higher head is triggered by head movement (for further discussion see also Gallego 2010). In my approach phase extension is indicated / made visible by head movement but defined independently of it (see (46) in Section 5 below).

A discussion of phase-hood in the sense of Chomsky (2000), however, is in order here. I find den Dikken's (2006) notion of an (inherent) phase as a predication domain more coherent than Chomsky's (2000) definition in terms of propositional object or full functional complex. Note also that Chomsky (2000) is denying phase status to TPs, even though TPs rather than vPs can be treated as representing (temporally specified) propositions that can be evaluated in terms of truth values. In the following, I will simply assume that the TP by (temporal) predication or as a propositional object counts as an (inherent) phase. This is implied by the bottleneck effect that we will discuss in detail in Section 4.1 below.

To simplify procedures and definitions, but maintaining the impact of (3), I will propose that movement may not go beyond a phase. Movement from phase to phase is possible by touching down in phase edges that count as mixed domains by belonging to the lower as well as to the higher phase, as is defined in (4). That is, an element from the v-domain, for instance, may move into the phase edge to the T-domain and further on into the T-domain, because this edge counts as belonging both to the v-domain and to the T-domain. Extending the phase edge of the C-domain thus means that the T-domain is extended up to ForceP and that the C-domain is eliminated resulting in a phase edge between the T-domain of the embedded clause and the v-domain of the matrix clause. I will come back to this point below. This means that the nature of a phase edge is defined dynamically by its context.

(4) Version of PIC: Movement may not cross phases. Phase edges permit movement into the higher phase since they belong both to the lower and the higher phase.

2.2 Privileged access in English

That subjects have privileged access to the C-domain can be shown with pertinent differences between local wh-movement of the subject and of the object in English, as is illustrated in (5). The data in (5) has first been discussed by Koopman (1983) and then later on under the heading of vacuous movement.[2] Note that wh-movement of the subject does not trigger *do*-support, while wh-movement of the object crucially does.[3]

(5) a. *Who met Mary at the party?*
b. **Who did meet Mary at the party?* (unless *did* is focused)
c. *Who did Mary meet at the party?*
d. **Who Mary met at the party?*

I assume that wh-words have a [+focus] and a [+wh] feature, the latter licensing interrogative force. In Hinterhölzl (2024) it is argued that subjects – exceptions aside and to be discussed below – move to [Spec,FinP] for reasons of referential anchoring (see also Section 4 below). If FinP counts as the phase-edge for movement into the C-domain, the wh-subject can be moved from [Spec,FinP] first to FocusP and then to ForceP to check its focus and wh-feature without any further ado. Wh-objects in English, on the other hand, cannot reach the C-domain to check their focus and wh-feature without any additional operation. Given that the phase-edge, namely [Spec,FinP], is occupied by the subject, they require an extension of the phase-edge to ForceP, which in V2-languages, as we will see below, is indicated by movement of the finite verb to the higher head.

For cases of long wh-movement, I assume that the embedded Force-head is not endowed with a [+wh-feature], but that the local focus-head is activated. Furthermore, I propose that constituents with unchecked features in the local domain can move to the phase-edge of that domain gratuitously,[4] in order to enter into a checking relation with a higher head endowed with the respective features.

For subject extraction – with FinP counting as the phase-edge – this means that the subject is stuck in the embedded FinP. It could move to FocusP to check its focus feature, but it cannot move any further to check its [+wh] feature. In particular, it cannot move to [Spec,ForceP], since the latter does not constitute the phase edge and it cannot move directly into the v-domain of the higher clause without violating locality in (4), deriving the *that-trace* effect in (6a).

---

[2] One may ask whether the wh-element in (5a) cannot be taken to remain in situ. I will not go into the details of this option and will assume without discussion that wh-movement to the C-domain in (5a) does in fact take place.
[3] Also, one may ask whether T to C movement is taking place in (5ac). I assume that in both cases T to Fin movement takes place. In (5c) also Fin to Force movement takes place, indicating phase extension. This implies that phase extension is dependent on a lexical exponent, otherwise (5d) should be grammatical and (5c) should be superfluous.
[4] An anonymous reviewer asks whether this movement is motivated by some edge feature. I prefer to treat it as a movement of last resort, but nothing much hinges on this decision.

(6) a. **Who did John say that met Mary at the party?*
b. *Who did John say that Mary met at the party?*

For object extraction the problem does not arise. Already for movement into the local FocusP, wh-movement of the object requires an extension of the phase edge from FinP to ForceP, hence the wh-object can move first to FocusP and then gratuitously into ForceP and since the latter counts as the overlapping area of two adjacent phase domain, namely the matrix v-domain and the embedded T-domain, it can move on into the v-domain in the matrix clause, accounting for the grammaticality of (6b).

In a nutshell the idea is simply this: subjects are privileged in that they can move into the C-domain in a more economic way than objects do. Wh-subjects can thus reach ForceP to check the interrogative force in main clauses without requiring a further operation of extending the phase domain. Wh-objects, and crucially also wh-adjuncts (not discussed in this paper), can do so only in a less economic way, requiring the extension of the phase domain to ForceP. For wh-subjects, however, we note that local economy leads to a problem when it comes to long distance movement.

How can this economy-induced problem with long distance movement of subjects be resolved? Note that the subject can be legally extracted from the embedded clause in its position in [Spec,FinP], if the entire C-domain is truncated, as is illustrated in (7a).

(7) a. $[_{\text{V-domain}}$ say $[_{\text{~~C-domain~~}}$ ~~that~~ $[_{\text{T-domain}}$ who met Mary at the party]]]
truncation phase edge
b. *Who did John say met Mary at the party*

In this case, the subject can move from the phase edge of the adjacent phase, namely the T-domain, into the v-domain in the matrix clause in one step. This 'truncation' operation necessarily deletes the complementizer in the C-domain, deriving the grammaticality of (7b).[5] But it also deletes the local Focus-head and the focus feature of the wh-subject remains unchecked in the embedded domain. But note that this generally does not lead to a problem given that also its [+wh] can remain unchecked in the embedded domain. I assume that both features are then checked in the matrix domain – with the wh-subject moving into [Spec,ForceP] to check its wh-feature and licensing interrogative force and with the insertion of an activated focus head in the matrix C-domain as an operation of last resort to check the focus feature of the lower subject. Note crucially that long wh-movement of the embedded subject triggers do-support in the matrix clause in (6b), as it does not move via [Spec,FinP] occupied by the matrix subject in the main clause.

[5] An anonymous reviewer asks how truncation actually works. One option would be that only the structure till FinP is projected or that the entire structure is first projected and structure above FinP is deleted in the course of the derivation. I prefer the first option so truncation should be understood metaphorically.

Note also that truncation of the C-domain has an effect similar to the extension of the phase edge in the C-domain to ForceP: it makes the embedded T-domain adjacent to the matrix v-domain and turns the phase edge of FinP into a mixed, lower T – higher v, phase category.

2.3 Privileged access in German

There is also evidence that the subject has privileged access to the C-domain in German. While almost any category – the subject, objects and adjuncts – can be topicalized to fulfill the strict V2-criterion of German, there is one common exception – namely, unstressed object pronouns – in difference to unstressed subject pronouns, which cannot occupy the clause-initial position, as is illustrated in (8). Object pronouns need to be stressed differently to subject pronouns and constitute either a focus constituent, as in (8d) indicated by a falling tone, or a contrastive or shifting aboutness topic, as in (8e) indicated by a rising tone, but cannot constitute an unstressed familiar topic, as in (8c) (cf. Frascarelli & Hinterhölzl 2007 for the notions of diverse topic types).

(8) a. *Er hat gestern die Maria besucht*
He has yesterday the-ACC Maria visited
'He has visited Mary yesterday'

b. *Es regnete gestern den ganzen Tag*
It rained yesterday all day-long
'It rained all day long yesterday'

c. **Sie hat der Nachbar gestern besucht*
Her has the-NOM neighbor yesterday visited
'The neighbor visited her yesterday'

d. *Sie (H+L) hat der Nachbar gestern besucht*
Her has the-NOM neighbor yesterday visited
'The neighbor visited HER (and not somebody else) yesterday'

e. *Sie (L+H) hat der Nachbar gestern besucht*
Her has the-NOM neighbor yesterday visited
'As for her, the neighbor visited her yesterday'

We noted in the previous section that subjects move into FinP to be referentially anchored. In general, I assume that contextual values – those pertaining to participants in the utterance situation (speaker and hearer) and those pertaining to discourse referents in the common ground – become accessible in FinP. All referential expressions, subjects and objects alike will thus enter into an Agree relation[6] with Fin^0 (for being assigned a referential value from the context and being licensed as familiar topics), but only the subject moves to [Spec,FinP] given that it normally constitutes the highest argument. Exceptions exist, if either the subject is not the highest argument (for example with certain psych-verbs) or if the subject is indefinite, as we will see below.

[6] The Agree relation is in terms of phi-features such that the most salient discourse referent with these phi-features is selected.

Let us assume (for the time being) that the V2 property is fulfilled by attracting the closest constituent to [Spec,ForceP], then if the speaker wants to utter a sentence with a familiar topic, they can use the subject as a clause-initial constituent but not the object. To take precedence over the subject, the object must take on a particular discourse role, either as a focus or as a special topic (a shifting aboutness topic or a contrastive topic). In this case, object movement into a specific position in the C-domain will trigger an extension of the phase edge to ForceP. Again, we see that subjects have a privileged access to the C-domain as opposed to objects, which does not come as a surprise, if we are right in treating privileged access as an economy effect.

There are cases where weak unstressed object pronouns can occupy the clause initial position,[7] as noted among others by Meinunger (2007). (9a) shows such an example that is due to Haider (2005). But what is crucial is that (9a) is possible as the subject is not referential, allowing the object to take precedence over the subject in entering in an Agree relation with Fin^0. As is illustrated in (9bc), as soon as a referential subject enters the game, the subject has to take precedence over the object. In other words, the data by Haider (2005) confirm the account of the observations in (8) in terms of referential anchoring of the arguments of the verb.

(9) *Dieses Schild können Sie genauso gut weglassen.*
This traffic sign can you as well remove.
a. *Es hat ohnehin keiner beachtet.*
It has anyway nobody noted.
b. *??Es hat er ohnehin nicht beachtet*
It has he anyway not noted
c. *Er hat es ohnehin nicht beachtet*
He has it anyway not noted

Now, wh-elements are treated as indefinites in traditional grammars. Thus, it is important to point out here that there is evidence that wh-elements have the extended structure of definite DPs. In Section 4.2 below, I will argue that definite DPs license an extra layer that hosts correlate elements that serve for referential anchoring, differently to indefinite ones. In Cimbrian, these correlate elements comprise adnominal *da* (here) and clitics. Wh-elements in Cimbrian can appear with adnominal *da* (cf. Cognola and Hinterhölzl 2020) and can be clitic-doubled in Spanish and Romanian (cf. Suner 1988), indicating an extended structure in their functional layer. Thus, wh-elements, like strong quantifiers, presuppose that a specific set of individuals is either discourse-given or uniquely identifiable in the reference situation. This interpretation of wh-elements is also supported by the fact that wh-subjects have privileged access to the C-domain in English, as shown in the previous section.

[7] I thank an anonymous reviewer for pointing these observations out to me.

2.4 Is there a *that-trace* effect in German?

The observation about privileged access of subjects in German is important, since there is a longstanding discussion of whether the *that-trace* effect exists in German. An experimental study by Featherstone (2005) shows that speakers of standard German are sensitive to the *that-trace* effect. In fact, it is well-known that north-German speakers do not like subject and object extraction from a *that*-clause, but use an embedded V2-structure instead, as is illustrated in (10).

(10) a. Wen sagte Hans [$_{\text{ForceP}}$ t hat [$_{\text{FinP}}$ der Nachbar [$_{\text{TP}}$ gestern t eingeladen]]]
Whom said John has the-NOM neighbor yesterday invited
'Who did John say that the neighbor has invited yesterday'

b. Wer sagte Hans [$_{\text{ForceP}}$ t hat [$_{\text{FinP}}$ t [$_{\text{TP}}$ gestern den Nachbarn eingeladen]]]
Who said John has yesterday the-ACC neighbor invited
'Who did John say has invited the neighbor yesterday'

According to our assumptions, V2 in (10) indicates that the phase-edge was extended from FinP to ForceP. The data in (10) are reminiscent of the strategy employed in Belfast English reported by Henry (1995) and discussed in Pesetsky & Torrego (2001), where object extraction makes use of an embedded V2 structure (11a). Differently from German, however, subject extraction in Belfast English does not make use of the same strategy, as is illustrated by the contrast between (11b) and (11c).

(11) a. *What did Mary claim* [t *did they steal* t] (Belfast English, Henry 1995, p. 128)
b. *Who do you think (*that) left*
c. **Who do you think did go to school* (Henry, p.c. to Pesetsky & Torrego)

In our approach, we can account for the data in (11) by noting that object extraction in Belfast English arguably makes use of phase-extension, but subject extraction makes the expected use of truncation. Note that in the present account, truncation and phase extension have the same effect: the phase edge between the embedded T-domain and the embedded C-domain turns into the phase edge between the embedded T-domain and the matrix v-domain, since phase extension extends the embedded T-domain up to the last head of the original C-domain, that is, Force^0.

The unexpected case so far is subject extraction in Standard German. If economy of subject movement into the C-domain also holds in Standard German, then the latter should also use the strategy of truncating the C-domain. I will argue in the following section that the German pattern in (10) follows from the fact that the phase-edge is regularly extended even for local movement of the subject via movement of the finite verb to ForceP due to the operation of a prosodic condition in the German C-domain that bleeds local economy (cf. the Prosodic Edge Condition in Section 3 below).

Let us now address the question of why the *that-trace* effect is disputed in German. The main reason is that South-German speakers, including myself, do not feel any difference between subject extraction and object extraction from a *that*-clause, as is illustrated in (12). Since the pre-eminent German linguists in the 1980s and 1990s were speakers of the dialect area of the South (Bavarian), like Haider, Bayer and Grewendorf, German was classified, on a par with Italian, as a language that is not subject to the *that-trace* effect.

(12) a. Wen sagte der Hans [$_{CP}$ dass [$_{IP}$ der Nachbar t eingeladen hat]]?
Whom said Hans that the-NOM neighbor invited has
'Who did John say that the neighbor has invited'

b. Wer sagte der Hans [$_{CP}$ dass t [$_{IP}$ t den Nachbarn eingeladen hat]]?
Who said Hans that the-ACC neighbor invited has
'Who did John say did invite the neighbor'

South German dialects differ from Standard German in that they trigger complementizer agreement, which is not visible in (12), but I will argue below that the presence of complementizer agreement can indicate that ForceP constitutes the phase edge of the embedded clause. We will come back to this issue when we discuss the French *que/qui*-rule and complementizer agreement in Sections 4 and 5 below.

The following section discusses which position the finite verb can be assumed to occupy in the extended C-domain. In principle, it could be any head-position in the left periphery. But I will assume that only two positions are eligible for the finite verb: Fin^0 and Force^0. The former is generally referred to as low V2 and the latter as high V2.

3 High and low V2

In this section, I will address the question of how to characterize V2 in the extended C-domain following Rizzi (1997). First of all, we have to distinguish strict linear V2 from systems where the finite verb moves into some functional head in the C-domain without strictly obeying the linear V2 rule.

Older Germanic languages, like OHG and OE, fail to respect a strict linear V2 constraint, since they allow for V1, V2 and V3 order in declarative clauses (cf. Hinterhölzl & Petrova 2010, Walkden 2015 for extensive discussion of this issue). While the V1-pattern in declarative clauses disappeared – apart from special usages in particular genres like jokes – in modern German, there are modern Germanic varieties that still exhibit V3-orders in declarative main clauses, most notably in West Flemish and in Kiezdeutsch (see below).

As is noted in Haegeman and Greco (2017), temporal adverbials in West Flemish, in difference to Standard Dutch, give rise to a V2 or a V3 pattern (cf. (13) below), with the crucial difference

that the V3 pattern does not allow for a reconstructed reading, implying that these adverbials are to be analysed as base-generated (high) topics. These adverbials are generally referred to as frame adverbials and comprise next to temporal adverbials also locative ones.

(13) a. *Oan-k toekwamen vielt den eletriek ut.*
when-1.SG arrived fell the electricity out

b. *Oan-k toekwamen, den eletriek viel ut.*
when-1.SG arrived, the electricity fell out
'When I arrived there was a power failure.'

In Dutch, a similar structure gives rise to a subject / non-subject asymmetry, as in (14) taken from Haegeman and Greco (2017): V3 order is possible with an initial object pronoun, but not with an initial subject pronoun. For a more extensive discussion of the contrasts between West Flemish and Dutch see Greco and Haegeman (2023).

(14) a. *Als er morgen een probleem is, mij moet je niet bellen.*
If there tomorrow a problem is, me must you not call
'If there is a problem tomorrow, don't call ME!'

b. **Als er morgen een probleem is, je moet mij niet bellen.*
If there tomorrow a problem is, you must me not call
'If there is a problem tomorrow, you don't have to call me!'

However, no such contrast appears with truly clause external adjuncts, that is, ones that modify the entire speech act, which arguably can be analysed as modifying ForceP, as in the German examples in (15ab). Here, a weak unstressed subject pronoun does not give rise to a marked status of the sentence. Standard German shows the same distinction as Dutch does, as illustrated in (16). As the contrast between (16b) and (16c) shows, the restriction does not apply to subjects in general, but only to weak unstressed subjects.

(15) a. *Wenn ihr Durst habt, es steht Bier im Kühlschrank*
If you thirst have it stands beer in-the fridge

b. *wenn es morgen ein Problem gibt, wen soll ich kontaktieren?*
if it tomorrow a problem exists, whom should I contact?
'Whom should I contact if there happens to be a problem tomorrow?'

(16) a. *wenn es morgen ein Problem gibt, MICH brauchst du nicht anzurufen*
if it tomorrow a problem exists, ME need you not call up
'You need not call me, if there should be a problem tomorrow'

b. **wenn es morgen ein Problem gibt, ich ruf dich an*
if it tomorrow a problem exists, I call you up
'If there is a problem tomorrow, I shall call you'

c. *wenn es morgen ein Problem gibt, PETER kann dir helfen*
if it tomorrow a problem exists, PETER can you help
'If there should be a problem tomorrow, Peter can help you'

Haegeman & Greco (2017) assume that all frame adverbials are base-generated outside of the clause and propose that there is a linking problem in (13b). I would like to propose that frame adverbials can attach both clause-internally and clause-externally, with the smaller structure being preferred for reasons of economy. Note that various authors that point out the occurrence of V3-structures in Older German (cf. Speyer 2008 for NHG and Petrova 2012 for Middle Low German) propose the structure in (17). Note, in particular, that they locate frame adverbials in the zone where base-generated topics are arguably licensed.

(17) [ForceP [FrameP [FocP [TopP [FinP [VP/IP …]]]]]] (Speyer 2008)

Furthermore, V3-structures appear in Kiezdeutsch, as is illustrated in (18) (cf. Wiese 2012, Sluckin 2025 for a more extensive discussion), raising the question how this variety differs from modern standard German. This is a tricky issue, since speakers of Kiezdeutsch do not make V2-'errors' with wh-movement or fronting of objects. Thus, they seem to have the German V2-system in place. I will thus argue below that their V2-system is akin to the standard German V2-system but that they fail to apply the prosodic condition in (20) below. Viewed in this way, V3-orders are direct indicators of a low V2-system and of the economy condition we argued for above.[8]

(18) *Heute ich geh Aldi* (Wiese 2012)
today I go Aldi (a type of supermarket)
'Today I will go to Aldi'

One option to explain the contrast between German and Dutch, on the one hand, and West Flemish (and Kiezdeutsch), on the other hand, would be to assume that while West Flemish allows for low V2, German and Dutch allow only for high V2. Note, however, that this assumption would leave us without an explanation for the occurrence of the *that-trace* effect in Standard German (and Standard Dutch). The proposal that I would like to make is the following: German allows for low V2 as well, but the prosodic condition in (20) triggers the extension of the phase edge to ForceP also with subject movement into the C-domain.[9]

In particular, I propose the following analysis for the contrast between a V2-structure and a V3-structure: If the verb raises to Force^0, a V2 structure will result and if the verb remains in Fin^0, a V3 structure can result (if a base-generated frame adverbial is present). This is illustrated in (21) below. Movement of the finite verb to Force^0 must then be taken to be triggered by weak subjects in Modern German and Modern Dutch, but is unnecessary in West Flemish, Kiezdeutsch and in various historical stages of German.

[8] There is one interesting issue: West Flemish speakers and speakers of Kiezdeutsch are not reported to produce V3 structures with subject questions. I will leave this issue for future research.

[9] I am following here an approach in which syntactic structure and prosodic structure are derived in a parallel and phase-based fashion (cf. Hinterhölzl 2019).

How can this be achieved? It is necessary to divide up the V2 properties in essentially two components: a syntactic part and an optional prosodic part. The syntactic part itself consists of two components, namely a requirement that Tense (pied-piping the finite verb) moves into the C-domain, and a constraint that limits the number of constituents that can be remerged in the C-domain to one (without requiring that at least one constituent be remerged in the C-domain). These conditions are summarized in (19).

(19) Compositional approach to V2:
   a) Tense moves to Fin^0 for licensing reasons, namely to be assigned values for the utterance situation and the reference situation
   b) The syntactic part of the V2 constraint is compatible with V1, V2 and V>2
   c) Strict linear V2 is enforced by the presence of a prosodic condition.

The basic idea is that Tense must enter in an Agree relation with Fin^0 to be assigned values for its two arguments, the utterance situation and the reference situation. I am following here the situation-based approach to Tense which assumes that Tense does not simply relate points in time or intervals, but expresses a relation between two situations (cf. Hinterhölzl 2024). This Agree relation can be expressed via head movement of T to Fin. I assume that this operation is universal, with V2 (meaning the finite verb appearing in the C-domain) resulting from the spell-out of the finite verb (or of the finite auxiliary) in the higher position.

Secondly, I propose that the famous bottleneck effect (cf. Haegeman 1996, Roberts 2004, Madaro, Tomaselli & Bidese 2025) can be reduced to the phase condition, assuming that (minimally) FinP counts as the phase-edge of the C-domain, thereby restricting movement from TP / IP to the C-domain to maximally one constituent. Finally, the prosodic part of the complex condition can be defined as given in (20). The condition in (20) re-enforces the phase edge nature of FinP. Phases are either typically used to locally constrain movement (as in den Dikken 2007) or to map spell-out domains onto prosodic constituents (as in Kratzer & Selkirk 2007). In this account, the phase edge in FinP serves both purposes: restricting movement and restricting spell-out options.

(20) Prosodic Edge Condition (PEC):
The finite verb must occupy a left-peripheral position in its prosodic phrase in the phase edge

With these conditions in place let us now come back to the issue of why V3 orders with base-generated topics do not appear in Standard German. (21ab) provide the essential syntactic structure of the sentence in (16b) above, while (22ab) indicate the relevant prosodic constituents in round brackets. Note in particular that (22a) violates the prosodic condition in (20): since arguments form a joint phonological phrase with the verb (cf. Gussenhoven 1992), the verb does not occupy a left-peripheral position in the phonological phrase containing it in the phase edge, when FinP counts as the phase edge of the clause

(21) a. [$_{\text{ForceP}}$ wenn es ein Problem gibt, [$_{\text{FinP}}$ ich ruf [$_{\text{TP}}$ dich an]]]
b. [$_{\text{ForceP}}$ wenn es ein Problem gibt, ruf [$_{\text{FinP}}$ ich [$_{\text{TP}}$ dich an]]

(22) a. [$_{\text{ForceP}}$ (Frame) [$_{\text{FinP}}$ ((Subj) V$_{\text{fin}}$).....]]
b. [$_{\text{ForceP}}$ (Frame) (V$_{\text{fin}}$ [$_{\text{FinP}}$ Subj) t]

What is the solution? The verb is raised to ForceP and the phase edge is extended to this phrase. As is illustrated in (22b), no violation of the prosodic edge condition in (20) arises here, since adjuncts and topics, familiar, shifting and contrastive Topics alike, are mapped onto a separate phonological phrase. This also holds for weak unstressed subjects: by moving on into the C-domain, they take on the role of either a familiar, contrastive or continuing aboutness topic and are thus phrased in a separate phonological phrase.

What are the implications of this account of the V2-phenomenon for the *that-trace* effect? We conclude that the phase edge is low in German, but is regularly extended to ForceP for objects and subjects alike (due to the presence of the prosodic requirement). We may thus assume that truncation of the C-domain is not an option in Standard German, hence a speaker of Standard German must use embedded V2-structures for the extraction of subjects and objects. What about southern German (Bavarian) dialects that generally accept subject extraction from a *that*-clause? This question will be addressed in the following section.

4 Subjects and complementizer agreement

In this section, I will prepare the grounds for the argument of why the *that-trace* effect does not appear in Southern German dialects,[10] which are dialects showing complementizer agreement and in general for the observation that the *that-trace* effect is bleeded, when the subject remains in a lower position. The core idea that I will pursue in the following sections is that if there is another element that undergoes movement to [Spec,FinP] instead of the subject, this will lead to the effect that subsequent movement of the subject, like movement of the object into the C-domain triggers the extension of the phase edge in the embedded clause to [ForceP].

[10] An anonymous reviewer pointed out that the systems of CA in Bavarian (much stronger) and in Alemennic dialects (less strongly realized) are not equal. An additional issue arises with South German standard speakers that do not use CA but nevertheless do not show *that-trace* effects. I would like to point out that a speaker even if not using CA to sound more like a speaker of standard German (second issue) or using a system with scarce evidence for CA (first issue) still has in place the (dialect) grammar in which the complementizer sits in the extended phase edge (as is required by the presence of CA). It is ultimately an empirical issue of what is going on when speakers speak two varieties. I am assuming that grammar change does not directly follow a change in language use. There could be an intermediate stage where the speaker assumes the presence of *pro* (note the pronominal origin of CA) before CA is taken to be lost altogether. But this question, important though it is, goes beyond the scope of this paper.

We will start out with discussing the interaction between wh-movement and subject licensing in Cimbrian, which will also provide us with an important clue for the correct analysis of complementizer agreement (CA) in Germanic. In particular, in this section I am re-assuming the data and arguments presented in Hinterhölzl (2024) to show that (referential) DPs are not interpreted by an assignment function in semantics but are licensed by entering into an Agree relation with a functional head in the C-domain. Furthermore, it is argued there that complementizer agreement is part of this licensing system. Thus, I will review here the most important data and arguments for this approach (for further details, the reader is referred to Hinterhölzl 2024).

4.1 Subjects and wh-movement in Cimbrian

In Cimbrian, as is illustrated in (23), there is a complementary distribution between preverbal subjects and the presence of a subject pronoun, or *da* ('here, there'), both of which are cliticized on to the verb (where *da* is spelled out as *-ta*). There is a large literature on the role of *da* in Cimbrian that I cannot do justice to in this paper.[11]

(23) a. *Bas *hat/hatta/hattar gekhoaft dar Luca in bottega*
what has/has-*da*/has-3.SG bought the Luca in shop
'What did Luca buy in the shop?'

b. *Dar Luca hat/*hatta/*hattar gekhoaft in liber in bottega*
Luca has/has-*da*/has-3.SG bought a book in shop
'Luca bought a book in the shop'

I will follow Cognola & Hinterhölzl (2020) in assuming that the pattern in (23) is due to referential anchoring of the subject. In particular, I propose that movement of a wh-element into the preverbal domain in (23) interferes with the licensing of subjects and that the presence of a low subject in (23a) provides direct evidence for the presence of a bottleneck effect in [Spec,FinP]. In the present account this means that [Spec,FinP] counts as the phase edge, explaining why the subject (in FinP) that arguably occupies an A-position can and does interfere with an operation of A'-movement.[12]

In other words, the wh-phrase *bas* in (23a) has to pass through [Spec,FinP] to reach [Spec,ForceP] to license the speech act of a question. Subjects in this case can be licensed in a lower position when they are doubled by *da* or by a subject clitic pronoun, as follows: if doubled by *da,* the subject is focused (new information or a contrastive focus), and if doubled by a pronoun, the

[11] The interested reader is referred to Bidese and Tomaselli (2005) and subsequent work; Kolmer (2005); Grewendorf and Poletto (2015) and Cognola and Hinterhölzl (2020).

[12] An anonymous reviewer pointed out that the C-domain is generally considered to be the target of A'-movement. Note, however, that in FinP we are dealing with a position in which (only) arguments are contextually valued.

subject is a topic. The idea that I would like to develop in the following section is that clitics and *da*, by undergoing head movement to Fin^0, serve to license the subject when the latter is unable to undergo movement to [Spec,FinP].[13]

4.2 The role of clitic pronouns and *da* in referential anchoring of arguments

In this section, I will argue that clitics and *da* serve to referentially anchor a subject in the context. In particular, I propose that definite DPs, in contradistinction to indefinite DPs, have an extra layer with an additional functional head that licenses a correlate DP in its specifier, as is illustrated in (24). Indefinite and weak (quantificational) DPs lack the respective layer and, thus must combine with an adverbial alternative anchor, as it occurs in the English existential *there* construction. I follow here a version of the Big DP-analysis of clitic doubling (cf. Torrego 1995, Uriagereka 1995), as in Cardinaletti (2019), in which the clitic does not constitute the head of an extended DP but is introduced in the specifier of a head external to DP. The evidence for this revised structure comes from participial agreement in Italian, which shows that clitics undergo XP-movement before they undergo head movement to their phonological host. In Cardinaletti's analysis, the head external to DP is K (for Case Phrase), but in view of the arguments of this paper, I propose that it is a referential head that guarantees co-evaluation between the associate and the DP. In particular, I propose that the DP-layer in (24) constitutes an island for extraction to explain why constituents internal to DP are subject to the Left Branch Condition, while the highest Specifier in the structure in (24) is not.[14]

(24) $[_{\text{RefP}}\ [_{\text{DP}}$ da/cl] Ref [D^0 $[_{\text{NP}}$ N]]]

Referential subjects can always anchor a predicate and obtain access to the value of their discourse antecedent, if they move into a pre-finite position, that is, into [Spec,FinP] in Cimbrian, as illustrated in (25). Furthermore, I propose that if a referential DP is moved into [Spec, FinP], no correlate is generated in [Spec,RefP] for reasons of economy, explaining the ungrammaticality of (25b).

---

[13] An anonymous reviewer pointed out that extraction of the correlate constitutes a violation of the Left Branch Condition. I like to point out that some cases of extraction from a left branch are allowed, as is indicated in (i). I will leave this issue for further research.

(i) a. Ich habe nicht davon gehört
I have not there-of heard
b. Da habe ich nicht [t von] gehört
There have I not of heard

[14] Note that the nature of the Left Branch Condition still needs to be explained and seems to be restricted to DPs, since in general there is no problem with extracting Specifiers from VPs, vPs, TP and CPs.

(25) a. *Dar Mario hat gekhoaft in liber*
Mario has bought a book

b. **Dar Mario hatta / hattar gekhoaft in liber*
Mario has-da / has-3.SG bought a book

Why is it that only subjects interact with clitic pronouns and *da* in Cimbrian? Must it not be assumed that all referential DPs have access to Fin^0? The latter assumption is correct. As stated above, I propose that Fin^0 enters into an Agree relation with all the referential arguments contained in TP and values them, but it will only attract the referential subject since it constitutes the highest argument in the structure. Given that it is the movement of the subject into [Spec,FinP] that interferes with the wh-movement via the bottleneck effect, it is the subject that interacts with the presence/absence of clitic pronouns and *da* in Cimbrian. When a subject is non-referential (i.e., indefinite), it will remain in a lower position and the sentence will be anchored via an alternative anchor, namely the reference situation argument of tense, as a thetic judgment.

To sum up what we have found so far, referential DPs cannot be interpreted within vP without any additional operation that connects them with the C-domain (see (23a) above).

Finally, let us take note here that Cimbrian does not make use of phase extension to allow for the movement of objects in the C-domain, but uses the vacant [Spec,FinP]. This analysis is indicated by the position of subjects, when the object undergoes wh-movement into the C-domain. While in German the subject follows directly the verb that has been raised to ForceP, the subject must stay in a lower position in Cimbrian, as is illustrated in (26ab).

(26) a. **Bas hat / hatta /hattar dar Luca gekoaft?* Cimbrian
what has/has-da/has-3.SG the Luca bought

b. *Bas hatta / hattar gekoaft dar Luca?* Cimbrian
what has-da/ has-3.SG bought Luca
'What did Luca buy?'

This strategy is arguably also employed in Italian and French main clauses: low subjects then go hand in hand with low V2 and a low phase edge. As is illustrated in (27) for Italian, the subject has to stay in a lower position to permit wh-movement of the object into the C-domain.

(27) a. **Che cosa ha Gianni comprato?* Italian
What thing has Gianni bought

b. **Che cosa Gianni ha comprato?*
What thing Gianni has bought

c. *Che cosa ha comprato Gianni?*
What thing has bought Gianni
'What did Gianni buy?'

To sum up, there is a second strategy next to phase extension to allow movement of the object into the C-domain in matrix clauses: Cimbrian, Italian and French have a more economical option of referentially anchoring the subject via head movement of a clitic correlate of the subject to Fin^0. In this way, [Spec,FinP] remains available for licit movement of the object into the C-domain without inducing an extension of the phase edge. As we will see in Section 5, this strategy is not available in embedded clauses in Italian and French for lack of movement of the finite verb into the C-domain, with head movement of the verb to Fin^0 being blocked by insertion of the complementizer. We turn now to the important question of the connection between subject licensing and complementizer agreement in West Germanic.

4.3 Complementizer Agreement in West Germanic

In this section, I will argue that complementizer agreement (CA) in West Germanic shows a pattern that is very similar to the facts of subject licensing in Cimbrian, hence I will argue that the phenomenon of CA is due to referentially anchoring the subject to the context. This section is based on van Koppen's work on CA in Dutch dialects and summarizes the discussion in Hinterhölzl (2024).

van Koppen (2006, 2012) argued that there are two types of CA to be distinguished. In a type A dialect, such as the dialect Tegelen Dutch, the agreement suffix is similar to the agreement suffix on a verb and CA is insensitive to subject movement and to subject modification (by focus). In a type B dialect, such as Hellendoorn Dutch, the agreement suffix (in C^0) differs from the agreement suffix on the verb, displaying the phenomenon of double agreement (DA). Furthermore, in these dialects, the agreement suffix is of pronominal origin and CA is sensitive to subject movement and subject modification. I will illustrate the variable nature of DA that occurs in various West Germanic dialects with van Koppen's (2012) data from Hellendoorn Dutch.

It is illustrated in (28) that if a subject is moved into a preverbal position, DA (*-e*) is ruled out, while if a subject stays in a lower position, as in a yes/no question that requires V1-order, DA is necessary. (29) illustrates that if a subject is modified by a focus particle, DA is ungrammatical, and (30) shows that in the presence of a frame adverbial, DA is excluded, while (31) shows that if a focused subject is moved to a higher position and has a definite reading, as is the case for the first person pronoun *wiej,* DA is again excluded.

(28) a. *Wiej binn-t /*binn-e den besten!*
we are-1.SG / are-DA the best!

b. *Binn-e /*binn-t wiej den besten?*
Are-DA /are-1.SG we the best?
'we are the best! / are we the best?'

(29) *dat / * darr-e [zölfs wiej] de wedstrijd wint.*
that / that-DA even we the game win
'that even we can win the competition'

(30) *dat / * darr-e [op den wärmsten dag van't joar] wiej tegen oonze wil ewärkt hebt.*
that / that-DA on the warmest day of the year we against our will worked have
'that we have worked against our will on the warmest day of the year'

(31) *WIEJ denkt Jan dat / *darr-e die pries ewönnen hebt, niet ZIEJ.*
we think Jan that / that-DA that prize won have, not they
'WE John thinks won that prize, not THEM.'
(Hellendoorn Dutch, van Koppen 2012:138)

In Hinterhölzl (2024), it is argued that the similar patterns of subject licensing in Cimbrian and CA in West Germanic call for a unified account, in which the facts of CA are explained in terms of anchoring the subject to the context. In particular, it is argued that in type B dialects, a clitic correlate of the subject moves into [Spec,FinP] to referentially anchor the subject to the context. In addition, it is proposed there that in type A dialects the clitic correlate has been reanalyzed as an Agreement marker that licenses the f-features of the *pro* correlate of the subject. These assumptions then predict that the subject in dialects / languages with CA does not move into [Spec,FinP] for referential anchoring. It will remain in a lower position and is thus eligible for phase extension, as objects are. This will be discussed in more detail in the following section.

5 Lower subjects and phase extension in embedded clauses

Let us take stock of what we have seen so far and what we generally know about the *that trace-* effect in diverse languages. In the previous section, we have seen that complementizer agreement allows for the subject in embedded clauses to remain in a position lower than [Spec,FinP]. Furthermore, Rizzi (1982) has argued that the *that-trace* effect is absent in Italian, since the subject can be extracted from a lower position in that language, as is illustrated in (33) below. Rizzi's argument is based on the observation that *ne*-cliticization is not possible from the preverbal subject position, but grammatical from postverbal constituents, objects and subjects alike, as is illustrated in (32).

(32) a. *Alcune (*ne) sono cadute nel mare.*
some CL are fallen into-the sea
'some (of them) have fallen into the sea'

b. *Gianni ne ha preso alcune.*
Gianni CL has taken some
'Gianni has taken some of it'

c. *ne sono cadute alcune nel mare.*
CL are fallen some into the sea
'Some of them have fallen into the sea'

Note first that the data in (32) are reminiscent of the behavior of subjects in Cimbrian. If the subject is moved into the C-domain, it alone can referentially anchor the predication, as is illustrated in (32a). If the subject remains in a lower position, the clitic correlate is moved into the C-domain, as is illustrated in (32c).

(33) a. **Quante hai detto che _ sono cadute?*
How many have-2.SG said that are fallen
'How many did you say that have fallen?'

b. *Quante hai detto che ne sono cadute _*
how many have-2.SG said that of them are fallen
'How many of them did you say that have fallen?'

(33a) clearly shows that Italian is sensitive to the *that-trace* effect. The ungrammaticality of (33a) follows in our account, given that *quante* arguably has moved into [Spec,FinP] in the embedded clause and is stuck there, assuming that Italian does not allow for truncation.[15] I will come back to this point below.

How can we then account for the grammaticality of (33b)? Note that phase edge evacuation, as it arguably occurs in main clauses in Italian, is not an option in embedded clauses, unless it is assumed that Italian allows for truncation of the C-domain with the complementizer being inserted (and remaining) in Fin^0. Note, however, that this is not an option given that there is strong evidence that the complementizer occupies a high position in the C-domain, preceding high (base-generated) topics. In (34) this topic is constituted by the frame adverbial *ieri*.

(34) *Che cosa ha detto Gianni che ieri ha dato a Maria, Luisa?*
What has said Gianni that yesterday has given to Maria, Luisa
'What did Gianni say that Luisa has given to Maria yesterday?'

The grammaticality of (33b), however, follows in our account assuming that phase extension is allowed in embedded clauses in Italian. In particular, note that if we assume that the clitic is moved into [Spec,FinP] in (33b) in embedded clauses, we expect that sub-extraction of the subject can proceed without problems via extending the phase edge.

[15] Here the question arises of whether it is arbitrary whether a language applies phase extension or truncation of the C-domain. This has to remain subject to further empirical work. For the time being, I note that languages that allow for low subjects (even if only in thetic judgments) use phase extension. If a subject is high in a language, as in English, extension is only possible for subextraction of the object, but not for the subject, rendering truncation of the C-domain as an operation of last resort.

In this respect, note first that sub-extraction of objects in Italian is possible with preverbal subjects, as illustrated in (35a). We have seen above that wh-movement of the object in main clauses requires a low (post-verbal) subject and does not make use of phase extension (due to the more economical option of phase edge evacuation via head movement of a clitic). The latter process obviously does not occur in embedded clauses in Italian. Thus, I propose the following solution: the complementizer in languages with low V2 is inserted in Fin^0 with the finite verb remaining in Tense. This implies that the correlate of the subject cannot reach FinP via movement of the finite verb to Fin^0, where the finite verb pied-pipes the relevant clitic. I therefore propose that the clitic correlate of the subject undergoes XP-movement to [Spec,FinP] creating the condition for triggering phase extension.

As in cases of V2, phase extension can then be indicated by movement of the complementizer from FinP to ForceP, as illustrated in (35b). Note, in particular, that phase extension and complementizer movement are necessary to bring the clitic in a valid licensing relation with its verbal host, for which I propose the following condition: clitic and verbal host need to be adjacent (and occupy the same phasal domain) to allow for the clitic to be phrased as part of the phonological word of the verb. Thus, the complementizer has to be moved into the higher phase edge.

(35) a. *Chi pensi che i linguisti hanno incontrato _?*
Who think-2.SG that the linguists have encountered
b. Quante hai detto [$_{ForceP}$ ~~quante~~ che [$_{FinP}$ ne ~~che~~ [$_{TP}$ sono cadute [~~ne [quante~~]]]]

Since standard Italian does not use subject clitics (36a), it is important for the present account to show for North Italian dialects that do have subject clitics, like Paduan, that a subject clitic is present and necessary with subject sub-extraction, as is illustrated in (36b).

(36) a. *Chi dice Gianni che ha visto Maria?* (Standard Italian)
Who says Gianni that has seen Maria
b. *Chi ze che dize Gianni ch'el gabia visto Maria?*
who is that says Gianni che-CL has seen Maria
(Paduan, p.c., Guglielmo Cinque)
'Who did Gianni say saw Maria'

Let us now move to the data that are known for French. French, differently from Italian, is taken to observe the *that-trace* effect, but differently also from English, it does not resort to truncation of the C-domain, but observes a variation of the complementizer that can be interpretated as displaying complementizer agreement, as illustrated in (37). Note, in particular, that French in contradistinction to Italian does not allow for postverbal definite subjects. Thus, I assume that the element showing up as a type of complementizer agreement in French is not an adnominal element as *da* in Cimbrian.

(37) a. **Qui a-t-il dit que _ voulait voir Marie?*
Who has-he said that wanted see Marie

b. *Qui a-t-il dit qui _ voulait voir Marie?*
Who has-he said that wanted see Marie
'Who did he say wanted to see Marie?'

Thus, I propose that *qui* is to be analysed as que+y, with *y* being the locative clitic, similar to *there* in English, rather than *da* in Cimbrian (cf. Taraldsen 1978 for proposing that *qui* should be analysed as *que+il*). I propose that *y* is inserted in [Spec,TP] and moved into [Spec,FinP] to referentially anchor the embedded clause as a thetic judgement (cf. Hinterhölzl 2019), allowing the wh-subject to remain in a lower position in the clause. The complementizer *que,* like in Italian, is thus taken to be inserted first in Fin^0 and then to be moved to Force^0 to indicate the extension of the phase edge, where the clitic is joined with the complementizer via local head movement, as illustrated in (38).

(38) Qui a-t-il dit [ForceP ~~Qui~~ que+y [FinP ~~y que~~ [TP ~~y~~ [AspP ~~Qui~~ voulait voir Marie]]]]?

These considerations bring us now to the difference of the *that-trace* effect in Standard German and Southern German. Southern German dialects, both Bavarian and Alemannic, do display CA. I propose that even when CA is not visible in cases like (39), the presence of CA in the variety indicates to native speakers that the complementizer in (39) occupies Force^0, testifying a high phase edge. In other words, I assume that a south German speaker when he drops CA to sound like a speaker of standard German, still has in place the dialect grammar, in which the complementizer sits in the extended phase edge. Native speakers of Standard German in our account must then be assumed to analyze the complementizer as occupying the lower phase edge, necessitating embedded V2-clauses for subject and object extraction.

(39) *Wer sagte Hans dass _ den Peter angerufen hat?* (Upper Austrian)
Who-NOM said John that _ the-ACC Peter called-up has
'Who did Hans say has called up Peter?'

The above discussion raises the question of what counts as evidence for the native speaker for the position of the complementizer in the C-domain. One possibility would be to assume that the complementizer in languages that do not observe V2 in declarative clauses is inserted in the highest head that is activated in the C-domain. This position would be FocusP for the complementizer in English embedded clauses containing a wh-word, hence the complementizer will be deleted under truncation of the C-domain. For languages with V2 in declarative clauses, I propose – again for reasons of economy – that the complementizer is assumed to occupy the lower Fin-head, unless there is direct empirical evidence (high topics or CA) that it occupies the higher Force-head. In addition, we have to assume that in languages that allow only low V2 in matrix questions, the complementizer is inserted in the Fin-head, but is regularly moved to the

Force-head to account for the Italian and French data. Note that the phase edge must always be taken by default to be FinP, with evidence for the complementizer in [ForceP] counting as direct indication of the extension of the inherent phase edge.

The presence of CA in a variety would then count as evidence for an exceptionally high position of the complementizer in that variety. Another piece of evidence comes from the presence of frame adverbials. It was pointed out that their presence bleeds the *that-trace* effect, as is illustrated for English in (40).

(40) *Who did Peter say that in this office _ is the boss?*

Remember also from the previous section that frame adverbials bleed CA, with CA-bleeding indicating either that the subject occupies a high position in the clausal domain, namely [Spec,FinP], or that there is an alternative (silent) anchor present in the structure. There are two options at our disposal now. A) We assume that the subject in (40) indeed occupies a high position before extraction and that it is the high position of the complementizer above the frame adverbial that indicates a high phase edge (and phase extension).[16] B) We assume that the subject in (41) occupies a low position before extraction and explain the absence of CA with the (silent) presence of another element that moves into [Spec,FinP] for referential anchoring instead of the subject. I will pursue option B) here, since it also allows for a more restrictive account of extending the phase edge.

Note first that frame adverbials serve to shift the reference situation to a new evaluation situation and are thus similar to shifting aboutness topics in the domain of individuals. This could indicate that the sentence is about a situation and that the reference situation argument of Tense, syntactically realized as *pro*, takes precedence over the subject in moving into FinP to enter into a local relationship with FrameP. In this case the subject would either be interpreted as weak definite (the unique set of people in this office in (40)) or be valued by Fin^0 without being subject to movement to FinP.

If this analysis of (40) is on the right track, we can assume that the extension of the phase edge is only possible (as a means of last resort) if there is an active head in the same phase (FocP in the embedded clause) that targets a constituent in a position lower than the phase edge, that is, lower than [Spec,FinP]. As illustrated in (41), this is not the case for subject extraction (SU), but relevant for the extraction of the direct object (DO).

(41) a. [V [$_{ForceP}$ [$_{FocusP}$ i-focus [$_{FinP}$ SU $_{[u\text{-}focus,\ uQ\ ]}$ that [$_{TP}$ ...]]]]] no phase extension
b. [V [$_{ForceP}$ [$_{FocusP}$ i-focus [$_{FinP}$ SU that [$_{TP}$ DO$_{[u\text{-}focus,\ uQ\ ]}$.....]]]]] phase extension

[16] Note that this is similar to Rizzi's (1997) original idea that Fin and Force are separate heads in the presence of topics, while they are collapsed otherwise.

I will finish this section with data reminding us that also low subjects in English lead to a bleeding of the *that-trace* effect, as is illustrated in (42a). This fact supports our assumptions in B) above: subjects give rise to the *that-trace* effect, since they are the prime candidates for referentially anchoring the sentence to the context.[17] If there is an alternative anchor – as in (42a) the deictic adverbial *there* – giving rise to a thetic judgment (cf. Hinterhölzl 2019), the *that trace*-effect does not arise. Additionally, it is important to note that subextraction of an alternative anchor does give rise to the *that-trace* effect, as is illustrated by the extracted locative in cases of locative inversion in (42b) taken from Bresnan (1977). In our account, this follows from the fact that the sentences with locative inversion constitute thetic judgments that are referentially anchored via movement of the locative PP into / through [Spec,FinP].

(42) a. *Which people did John say that there were _ dancing in the street?*
b. *In which villages do you believe (*that) are found the best examples of this cuisine?*

6 In conclusion: Evaluating the Nature of the *that-trace* effect

Cowart & McDaniel (2021) point out that depending on the theoretical framework, the *that-trace* effect is viewed either as a grammatical violation or as a by-product of sentence processing. They also note that the phenomenon remains a conundrum, since it appears in a number of unrelated languages, but also varies in closely related languages (Cowart & McDaniels 2021: 258). Haider (2023) points out that the *that-trace* effect is not categorical but gradual – typical for processing effects – by citing the acceptability scores of Sobin (2002) of the examples in (43).

(43) a. *Who did you say that saw Elvis last night.* 64% good, 27% possible, 9% bad
b. *Who do they think that might visit the pope.* 22% good, 23% possible, 56% bad

First, I would like to point out that Sobin (2002) himself admits that the number of subjects in his experiment, which also comprised only a couple of sentences, is very low. Second, having a category 'possible' in a grammaticality judgment is not very helpful.[18] Third, I would like to note that the gradient nature of the data does not necessarily reflect the absence of a categorical grammatical condition, but rather indicates the variable evaluation of the presence and the position of the complementizer. Note that the presence of the complementizer in subject extractions in English does not necessarily lead to ungrammaticality, as is indicated in (40) and (42a) above. In the present account, the judgement of the speaker crucially depends on whether there is sufficient evidence for a high position of the complementizer in the relevant example. Note in this respect that (43a) involves a bridge verb that allows for embedded root phenomena (embedded

[17] A similar conclusion is reached in Bayer and Salzmann (2013).
[18] This is so, since it is not entirely clear what subjects mean when they judge a sentence as possible. Is it possible, since they would not use the sentence but have heard other speakers using it? Does 'possible' mean that the sentence is not perfect but anyway grammatical? In general, it will induce a speaker not to make a decision between grammatical or ungrammatical in case a sentence for whatever reason sounds weird to them.

V2, a dependent assertion). This could indicate to the native speaker that ForceP is activated and that the complementizer occupies a high position in the embedded clause, presenting evidence for phase extension. However, this is not categorical because the bridge verb may also be taken to simply embed a proposition. Similar considerations apply to (43b), where there is less evidence for a high complementizer, hence the expected lower ratings.

Sobin (2002) provides also a common account of the *that-trace* effect and the deletion or non-deletion of the complementizer in subject and object relatives. Note first that a *that-trace* effect is not expected in subject relatives in the present account, since the subject is not actually extracted from the embedded clause – but only moved into the local C-domain in standard accounts of relativization as well as in more recent matching approaches (cf. Cinque 2014). Thus, I argue that what happens in relative clauses is not a truncation of the C-domain, but the optional deletion of the complementizer that indeed is subject to processing constraints, as is illustrated in (44). While the correct interpretation of the initial constituent can be resolved locally in object relatives – after processing *Mary met* in (44a) – the interpretation of the initial constituent must be re-elaborated and can only be corrected after processing the entire clause, leading to a severe processing difficulty. Nothing like this is going on in Sobin's sentences in (43): they are easy to process, but may differ to various degrees in their evaluation of whether they represent the most economic version of expressing the relevant thought.

(44) a. *The woman Mary met yesterday has just called*
b. **The woman met Mary yesterday has just called*

Furthermore, subject relatives indicate that the notion *that-trace* effect is rather a misnomer. There is nothing wrong with the sequence *that trace* in grammar. Rather, as I have argued, it arises as an economy effect in the extraction from the embedded clause. I consider this a great advantage of the present account over past accounts.

To conclude, the present paper has placed the *that-trace* effect in a context of interconnected phenomena of (referentially) anchoring the sentence to the context in which the subject (as the highest argument) stands out. A maximally simple account of the *that-trace* effect was proposed that links it to a general condition of derivational economy. We have pointed out a number of repair strategies, namely truncation, phase edge extension and complementizer agreement that explain the appearance or non-appearance of the *that-trace* effect in German, English, French and Italian.

Furthermore, I have presented a new concept of phase-based movement, namely flexible phase edges that can be extended for syntactic reasons, but also due to prosodic conditions. Furthermore, I have presented empirical evidence for the bottleneck effect in Cimbrian, Italian and French, explaining why wh-movement of objects is necessarily connected to a low position of subjects. Low subjects in wh-questions are not required in high V2 languages like German and Dutch, since they regularly take recourse to phase extension.

References

Bayer, Josef & Martin Salzmann. 2013. *That*-trace effects and resumption – How improper Movement can be repaired. In P. Brandt & E. Fuß (eds.), *Repairs. The added value of being wrong*, 275–333. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614510796.275.

Bidese, Ermenegildo & Alessandra Tomaselli. 2005. Formen der 'Herausstellung' und Verlust der V2-Restriktion in der Geschichte der zimbrischen Sprache. In Ermenegildo Bidese, James R. Dow and Thomas Stolz. Bochum (eds.), *Das Zimbrische zwischen Germanisch und Romanisch*, 71–92. Brockmeyer.

Bidese Ermengildo & Alessandra Tomaselli. 2018. Developing pro-drop: The case of Cimbrian. In Federica Cognola & Jan Casalicchio (eds.), *Null Subjects in Generative Grammar*, 52–69. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198815853.003.0003.

Bresnan, Joan. 1977. Variables in the theory of transformations. In Peter Culicover, Thomas Wasow and Adrian Akmajian (eds.), *Formal Syntax*, 157–196. New York: Academic Press.

Cardinaletti, Anna. 2019. Cliticization as extraction: The big DP hypothesis revisited. *Revista da Associão Portuguesa Linguìstica*. 1–16. https://doi.org/10.26334/2183-9077/rapln5ano2019a1.

Chomsky, Noam. 2000. Minimal inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays in minimalist syntax in honour of Howard Lasnik*, 89–115. Cambridge, MA: MIT Press.

Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8. 425–504.

Cinque, Guglielmo. 2014. *The syntax of relative clauses*. Cambridge: Cambridge University Press.

Cognola, Federica & Roland Hinterhölzl. 2020. Syntactic and semantic restrictions in the licensing of subjects in Cimbrian main clauses. *Linguistische Berichte*. https://doi.org/10.46771/2366077500263_2.

Cowart, Wayne & Dana McDaniel. 2021. The that-trace effect. In *The Cambridge handbook of experimental syntax*, 258–277. Cambridge: Cambridge University Press.

den Dikken, Marcel. 2007. Phase Extension: Contours of a theory of the role of head movement in phrasal extraction.*Theoretical Linguistics* 33(1). Walter de Gruyter. https://doi.org/10.1515/TL.2007.001.

Featherstone, Sam. 2005. *That*-trace in German. *Lingua* 115. 1277–1302. https://doi.org/10.1016/j.lingua.2004.04.001.

Frascarelli, Mara & Roland Hinterhölzl. 2007. 'Types of Topics in German and Italian'. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 87–116. Amsterdam: Benjamins. https://doi.org/10.1075/la.100.07fra.

Gallego, Angel. 2010. *Phase theory.* Linguistics Today. Amsterdam: Benjamins. https://doi.org/10.1075/la.152.

Greco, Ciro & Liliane Haegeman. 2023. Frame setters and micro-variation of subjects. In *Adverbial resumption in verb second languages,* 43–77. Oxford University Press. https://doi.org/10.1093/oso/9780197651148.003.0003.

Grewendorf, Günther & Cecilia Poletto. 2015. Relative clauses in Cimbrian. In Di Domenico, Elisa et al. (eds.), *Structure, Strategies and Beyond. Studies in Honor of Adriana Belletti*, 393–416. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/la.223.18gre.

Gussenhoven, Carlos. 1992. Sentence accent and argument structure. In I. Roca (ed.), *Thematic structure: Its role in grammar*, Foris. 79–106. Berlin & New York.

Haegeman, Liliane. 1996. Verb second, the split CP, and null subjects in early Dutch finite clauses. *Geneva Generative Papers* 4. 135–175.

Haegeman, Liliane & Ciro Greco. 2017. Main clause external constituents and the derivation of subject initial verb second. Ms. University of Ghent.

Haider, Hubert. 2005. How to turn German into Icelandic – and derive the VO-OV contrasts. *Journal of Comparative Germanic Linguistics* 8. 1–53. https://doi.org/10.1007/s10828-004-0293-0.

Haider, Hubert. 2023. Is ChatGPT a grammatically competent informant? lingbuzz/007285.

Henry Alison. 1995. *Belfast English and Standard English.* Oxford: Oxford University Press.

Hinterhölzl, Roland. 2019. Subjects, topics and anchoring to the context. *Syntax* 22(2–3, Special Issue). 199–228. https://doi.org/10.1111/synt.12179.

Hinterhölzl, Roland. 2021. V2 and topicalization in Germanic and Romance. In Sam Wolfe & Christine Meklenborg (eds.), *Continuity and variatin in Germanic and Romance*, 195–222. Oxford: Oxford University Press.

Hinterhölzl, Roland. 2024. Complementizer agreement and the licensing of DPs: An account in terms of Referential Anchoring. *Languages* 9. 1–14. https://doi.org/10.3390/languages9020049.

Hinterhölzl, Roland & Svetlana Petrova. 2010. From V1 to V2 in West Germanic. *Lingua* 120. 315–328. https://doi.org/10.1016/j.lingua.2008.10.007.

Kandibowicz, Jason. 2006. Comp-trace effects explained away. In Donald Baume, David Montero and Michael Scanlon (eds.), *Proceedings of 25th WCFL*, 220–228. Somerville, MA: Cascadilla Proceedings Project.

Kolmer, Agnes. 2005. L'espletivo *da* come espletivo della posizione del soggetto enclitico pronominale nel Cimbro di Luserna. In Walter Breu (ed.), *L'influsso dell'italiano sulla grammatica delle lingue minoritarie. Problemi di morfologia e sintassi*, 55–92. Rende: Centro Editoriale e Librario Università della Calabria.

Koopman, Hilda. 1983. ECP-effects in main clauses. *Linguistic Inquiry* 14. 346–350.

Kratzer, Angelika & Elisabeth Selkirk. 2007. Phase theory and prosodic spell-out: the case of verbs. *The Linguistic Review* 29(2–3). 93–155.

Madaro Romano, Alessandra Tomaselli & Ermenegildo Bidese. 2025. Deriving CP-expansion in the German enclave varieties across Northeast Italy: Cracks in the bottleneck? In Sarah Harchaoui & Pierre-Yves Modicom (eds.), *Verb-third phenomena in Germanic verb-second languages: Historical and variational perspectives*, 291–344. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.17077198.

Meinunger, André. 2007. About object es in the German Vorfeld. *Linguistic Inquiry* 38(3). 553–363. https://doi.org/10.1162/ling.2007.38.3.553.

Padovan, Andrea, Ermenegildo Bidese & Alessandra Tomaselli. 2021. Circumventing the 'that-trace' effect: Different strategies in Germanic and Romance. *Languages* 6(2). 84. https://doi.org/10.3390/languages6020084.

Pesetsky, David. 2017. Complementizer-trace effects. In Martin Everaert & Henk van Riemsdijk (eds.), *Companion to Syntax*, 2nd edition. Wiley, Blackwell.

Pesetsky, David & Esther Torrego. 2001. T-to-C movement: Causes and consequences. In Michael Kenstowics (ed), *Ken Hale: A life in language*, 355–426. Cambridge, Mass.: MIT Press.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of Grammar*, 281–337. Dordrecht: Kluwer.

Roberts, Ian G. 2004. The C-system in Brythonic Celtic languages, V2, and the EPP. In Luigi Rizzi (ed.), *The structure of CP and IP*, 297–328. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780195159486.003.0010.

Sluckin, Benjamin. 2025. Revisiting the syntax and the development of Kiezdeutsch V3: A new perspective. *Journal of Germanic Linguistics* 37(1). 64–124. https://doi.org/10.1017/S1470542724000114.

Sobin, Nicholas. 2002. The comp-trace effect, the adverb effect and minimal CP. *Journal of Linguistics* 38(3). 527–560. https://doi.org/10.1017/S0022226702001652.

Speyer, Augustin. 2008. Doppelte Vorfeldbesetzung im heutigen Deutsch und im Frühneuhochdeutschen. *Linguistische Berichte* 216. 457–487. https://doi.org/10.46771/2366077500216_3.

Suner, Margarita. 1988. The role of Agreement in Clitic-doubled constructions. *Natural Language and Linguistic Theory* 6. 391–434.

Taraldsen, Knut Tarald. 1978. On the nominative island condition, vacuous application and the that-trace filter. Unpublished manuscript. Indiana University Linguistics Club.

Torrego, Esther. 1995. On the nature of clitic doubling. In Hector Campos & Paula Kempschinsky (eds.), *Evolution and revolution in linguistic theory*, 399–418. Washington, DC: Georgetown University Press.

Uriagereka, Juan. 1995. Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry* 26. 79–124.

van Koppen, Marjo. 2006. Complementizer agreement. Master's thesis, University of Utrecht, Utrecht, The Netherlands

van Koppen, Marjo. 2012. The distribution of phi-features in pronouns. *Natural Language and Linguistic Theory* 30. 135–77. https://doi.org/10.1007/s11049-011-9159-8.

Vikner, Sten, Ken R. Christensen & Anne N. Nyvad. 2017. V2 and cP/CP. In Laura R. Bailey & Michelle Sheehan (eds.), *Order and structure in syntax I: Word order and syntactic structure*, 313–324. Berlin: Language Science Press.

Walkden, George. 2015. Verb-third in early West Germanic: A comparative perspective. In Theresa Biberauer & George Walkden (eds.), *Syntax over time*, 236–248. Oxford: Oxford University Press.

Wiese, Heike. 2012. *Kiezdeutsch: ein neuer Dialekt entsteht*. München: C.H. Beck.

ZS ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 2
https://doi.org/10.18148/zs/2026-2002

Hengzhi Hu, National University of Malaysia
Luying Zheng, National University of Malaysia, p119571@siswa.ukm.edu.my

Language Change: Theoretical and Empirical Perspectives

Reviewed Publication: Elitzur A Bar-Asher Siegal, Nora Boneh, Eitan Grossman & Aynat Rubinstein (eds.). 2025. *Language change: Theoretical and empirical perspectives*. Cham: Springer.

Research on language change sits at a durable crossroads between description and explanation. We can now document change with increasingly fine-grained corpora and typology, yet persistent debates remain about what counts as evidence (especially for meaning change), how to separate semantic change from pragmatic enrichment, and which mechanisms (e.g., reanalysis, inference, acquisition, diffusion, communicative pressures) are actually doing the explanatory work (Millar 2024). These questions matter directly for journals whose scope includes language change, because such venues serve a cross-subfield readership that expects both empirical credibility and theoretical cumulativeness. *Language Change: Theoretical and Empirical Perspectives*, edited by Bar-Asher Siegal et al. (2025a), responds to this landscape with a clear intervention: progress on language change, particularly meaning-related change, requires explicit theorizing about what the phenomenon is, how it can be detected in historical data, and what predictions a proposed mechanism makes. The volume concludes an interdisciplinary research program on diachrony and formal semantics, which sought to foster sustained collaboration between historical linguistics and formal semantics rather than parallel lines of inquiry.

The book's fifteen chapters span negation, cyclicity, aspect, modality, information structure, and related interface problems. Chapter 1 sets the conceptual agenda for "formal diachronic semantics." Chapter 2 offers a diachronic account of why newly grammaticalized future markers often pattern asymmetrically under negation; it argues that these asymmetries follow from

Cite as: Hu, Hengzhi & Luying Zheng. 2026. Language Change: Theoretical and Empirical Perspectives: Reviewed Publication: Elitzur A Bar-Asher Siegal, Nora Boneh, Eitan Grossman & Aynat Rubinstein (eds.). 2025. *Language change: Theoretical and empirical perspectives*. Cham: Springer. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi. org/10.18148/zs/2026-2002
Received: 07 January 2026; **accepted:** 02 February 2026; **published:** 08 June 2026
Published by: publikon / dgfs (Linguistic Society of Germany)

how emergent futurity interacts with modality and scope during grammaticalization. Chapter 3 develops a typology/history of "two types of negation"; it treats the two types as semantically distinct classes whose differing distributions help explain recurrent trajectories in negative systems. Chapter 4 argues for cyclicity between anaphors and intensifiers; it shows that recurrent reanalyses between these functions can be modeled without abandoning constraints on binding and emphasis-related meaning. Chapter 5 traces the auxiliation of Hebrew *come*. Chapter 6 treats Latin *quidem* in contrast/negation environments; it uses these environments as diagnostics to sharpen *quidem*'s discourse–semantic contribution at the focus/contrast–negation interface.

The remaining chapters extend this program across domains and methods. Chapter 7 explains preterit loss in Southern German via interacting intra- and extra-linguistic factors. Chapter 8 theorizes "innovation" in grammatical morphology. Chapter 9 develops cyclicity effects in presupposition development (including a diachronic adaptation of "Maximize Presupposition"); it argues that competition among alternatives can drive historical shifts in presuppositional strength and distribution. Chapter 10 discusses a stative-to-perfective pathway in Semitic in typological perspective. Chapter 11 reconstructs the diachronic semantics of two Hungarian "almost" items; it shows how near-synonyms can split into stable semantic niches as scalar and counterfactual components become differently conventionalized. Chapter 12 links subjectivity with shifts in subordination/coordination at the semantics–syntax interface in Hebrew; it argues that changes in speaker-oriented meaning are coupled with changes in clause-linkage strategies, locating subjectification at a concrete interface locus. Chapter 13 analyzes Biblical Hebrew cardinal constructions with ordinal meaning using synchronic variation to infer diachronic developments. Chapter 14 uses modality pathways as a test case for mechanisms of meaning change. Chapter 15 reports an artificial-language learning experiment on the emergence of differential object marking from atypical information structure; it provides experimental support that givenness-driven discourse pressures can motivate the conventionalization of a new morphosyntactic contrast over time.

For language-change scholarship, the volume's core strength is that it consistently treats explanation as more than narrative plausibility. The editors foreground problems that historical linguists often acknowledge but do not always operationalize: identifying semantic change when older and newer uses overlap; distinguishing semantic change from pragmatic enrichment; and deciding whether a recurring "pathway" reflects a mechanism or merely a descriptive generalization (Bar-Asher Siegal et al. 2025b). Many chapters respond by stating diagnostics and predictions rather than relying on umbrella labels, most notably Chapters 2, 3, 4, 6, 9, 11, 12, and 15. That style matters for a language-change readership because it makes disagreement productive: when a mechanism is formulated as a prediction about distribution, competition between forms, or compositional contribution, debate can advance with sharper evidence rather than terminological substitution.

A second strength is its coherence around recurring concerns—negation/polarity asymmetries, cyclicity, and interface-driven pathways (Millar 2024)—without enforcing methodological uniformity. The negation-focused contributions, for instance, treat asymmetries as constraints on how innovations diffuse and stabilize, not as language-particular curiosities. The "cycle" motif likewise appears in different domains (anaphor–intensifier relations; presupposition marking), inviting a broader question: are cycles best understood as properties of specific subsystems, or as outcomes of tensions between inference, economy, and form–meaning alignment? The collection's breadth prevents formal diachronic semantics from collapsing into a single favored mechanism; instead, it argues that explanation benefits from explicitness, regardless of domain.

Methodologically, Chapter 15 by Tal et al. (2025) signals an especially consequential expansion of what can count as evidence in language-change argumentation. It tests a specific hypothesis about the emergence of differential object marking: that differential object marking may originate as a response to atypical information structure (givenness) and only later be re-analyzed as semantically conditioned (e.g., animacy). The reported outcome is important for language-change theory precisely because it is not simply confirmatory: while the experiment supports the emergence of differential object marking under atypical information structure, this effect did not generalize to animate objects as such, despite the input correlation between animacy and givenness. In production, participants did not increase marking for animate objects overall (and on some testing days the trend even went in the opposite direction), and the statistical model reports no reliable main effect of animacy on object marking. This is a methodological service to the field. It shows how experiments can test the plausibility of proposed micro-mechanisms without recreating historical ecologies, and it pressures pathway narratives to specify which steps are expected under learnability and which require additional conditions (contact, register stratification, sociolinguistic diffusion, acquisition setting). For readers invested in language change as an explanatory science (Shalini & Perono Cacciafoco 2023), this chapter reframes experimentation as a constraint on mechanism claims rather than as an alternative to historical data.

Despite its broad title, the volume's center of gravity is semantic and morphosyntactic, consistent with its origin in a program on diachrony and formal semantics; readers seeking sustained engagement with phonological change, contact-induced change as the primary driver, or variationist quantitative modeling will therefore find those dimensions largely peripheral. In addition, the language coverage is uneven, with Hebrew/Semitic data anchoring multiple chapters (auxiliarization, subordination/subjectivity, numerals, modality pathways, stative-to-perfective), which is a major asset for Semitic diachrony but may feel narrower than the title implies for readers who expect typological balance across unrelated families. Finally, accessibility varies: some chapters are readily usable by historical linguists without extensive formal-semantics training, while others presuppose familiarity with formal semantic argumentation, and the absence of

a brief concluding synthesis means that cross-chapter generalizations (e.g., when semantic vs pragmatic explanations diverge most sharply) are left for the reader to assemble.

The book's audience is best described in language-change terms. It will be most valuable to advanced graduate students and researchers working on meaning change, grammaticalization pathways, and interface phenomena, especially those who want to connect diachronic evidence to explicit semantic hypotheses and debates about mechanisms and recurrence. Formal semanticists interested in diachrony will find historically grounded problems that can discipline synchronic theory-building, while semiticists and scholars of Hebrew diachrony will find a rich cluster of theory-forward studies. For language-change scholars outside semantics, the book is best approached less as a survey and more as a methodological demonstration: it shows how to raise the evidential bar for claims about pathways and mechanisms by making analyses explicit, testable, and easier to contest.

Overall, the book is a strong contribution to language-change scholarship because it insists that meaning-related change can be analyzed with the same explanatory ambition and the same demand for constraints that historical linguistics brings to other domains. Its most persuasive claim is methodological: language-change research becomes more cumulative when "theoretical" and "empirical" are treated as mutual checks implemented through diagnostics, typological comparison, corpus reasoning, and (where appropriate) experimental constraints.

References

Bar-Asher Siegal, Elitzur A., Nora Boneh, Eitan Grossman & Aynat Rubinstein (eds.). 2025a. *Language change: Theoretical and empirical perspectives*. Cham: Springer. https://doi.org/10.1007/978-3-031-85292-3

Bar-Asher Siegal, Elitzur A., Nora Boneh, Eitan Grossman & Aynat Rubinstein. 2025b. Introductory notes to language change: Theoretical and empirical perspectives. In Bar-Asher Siegal, Elitzur A., Nora Boneh, Eitan Grossman & Aynat Rubinstein (eds.), *Language change: Theoretical and empirical perspectives*, 1–14. Cham: Springer. https://doi.org/10.1007/978-3-031-85292-3_1

Millar, Robert McColl. 2024. *Linguistic contact and language change: An introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009071093

Shalini, V & Francesco Perono Cacciafoco. 2023. The evolutionary origins of language: Validating a new theory. *Annals of the University of Craiova. Series Philology. Linguistics* 45(1–2). 403–460. https://doi.org/10.52846/aucssflingv.v45i1-2.115

Tal, Shira, Kenny Smith, Jennifer Culbertson, Eitan Grossman & Inbal Arnon. 2025. From atypical information structure to semantically-conditioned differential object marking: An experimental study. In Bar-Asher Siegal, Elitzur A., Nora Boneh, Eitan Grossman & Aynat Rubinstein (eds.), *Language change: Theoretical and empirical perspectives*, 433–459. Cham: Springer. https://doi.org/10.1007/978-3-031-85292-3_15