Variation and dynamics of “complementizer agreement” in German
Analyses from the Austrian language area

Matthias Fingerhuth and Alexandra N. Lenz
University of Vienna

To date, there has been limited empirical research on complementizer agreement (CA). We investigate CA drawing on a corpus of 144 speakers from 13 locations across Austria that was elicited through computer supported language production experiments and recorded in conversations. We investigate the linguistic factors that govern the occurrence of CA, as well as its areal distribution. We further explore the role of CA in speakers’ linguistic repertoires. The study finds evidence for the hypothesis that the (non-)occurrence of CA is strongly dependent on the structure of its hosting C-elements and finds regional patterns. It also identifies CA as a phenomenon which speakers place in a non-standard register. We use the collected data to test a theory of the emergence of CA from clitic pronouns for Bavarian varieties of German.

Keywords: complementizer agreement, morphology, syntax, grammaticalization, dialectology, linguistic repertoires, methodology, experimental linguistics

1. Introduction

This contribution focuses on complementizer agreement (CA), a morphosyntactic phenomenon which involves an agreement morpheme attaching in the right-hand side of the C-domain of a subordinated clause. This typologically rare phenomenon appears in Continental West Germanic languages or their non-standard varieties respectively (Weiß 2005, 2018; de Vogelaer & van der Auwera 2010), but has also been described for e.g. Bantu (Diercks 2013; see Bohn & Weiß 2016:160 fn. 2, for comments that question the classification as CA). This phenomenon is of interest to a broad range of linguistic subdisciplines, including syntactic theory, areal linguistics, typology, as well as sociolinguistics. Regarding
German-speaking areas, the majority of research has focused on the Bavarian dialect area, which extends over three European countries: South-Eastern Germany, most of Austria, and South Tyrol in Northern Italy (see map 1 in Lenz et al. 2014: 25). The existing research primarily considers the phenomenon due to its relevance to syntactic theory and has only rarely explored the phenomenon from a variationist perspective, on a broad empirical basis. This article aims to address this desideratum. Based on data collected across Austria, this article will consider the following questions:

– Which linguistic factors influence CA? How does CA occur dependent on number and person of the finite verb, the complexity of subordinate clause introductions, and the verbal and pronominal morphology of the varieties in question?
– Which areal distribution does CA show in the dialects of Austria? Which areal-horizontal structures appear?
– Where can we locate CA on the social-vertical spectrum on the dialect-standard axis of German?

Building on the data, we will discuss how these findings match a theory of the emergence of CA as proposed by (Weiß 2018). Further, the article contributes to linguistic methodology by introducing an “experiment” which is designed to investigate CA under the consideration of potential factors that govern its occurrence. We want to demonstrate the effectiveness of this innovative elicitation method and encourage the use of similar experiments in research on syntax. Following this introduction, the article is structured as follows: Chapter 2 provides an overview of the current state of research on CA, with a focus on German. This includes theory-oriented approaches from syntactic research, as well as descriptive accounts from German dialectology and a discussion of a theory of the emergence of CA. Chapter 3 provides the analysis of our data following a description of the underlying corpus. This corpus consists of data collected using different research methodologies and is currently created by the Special Research Programme (SFB) on “German in Austria. Variation – Contact – Perception” (DiÖ). This study draws on data from language production experiments (see Lenz et al. 2019), as well as data from open conversations. The findings are used to test a theory of the emergence of CA from clitic pronouns. The article ends with a summary of the findings to the previously mentioned questions and provides an outlook on further work.
2. Research on complementizer agreement in German

2.1 State of research

In a long history of description in German and Dutch dialectological traditions, several names have been used to describe CA. It has been labeled as “Flexion der Konjunktion” (‘inflection of the conjunction’, Weise 1907), “congruerende voegwoorden” (‘inflecting conjunctions’, van Haeringen 1939), “Conjugation des Bindewortes” (‘conjugation of the linking word; Schiepek 1899: 76), “Suffigierung der Personalpronomina” (‘suffixation of the personal pronouns; Pfalz 1918) or “personmarkierte Einleitung von Nebensätzen” (‘introduction of embedded clauses marked for person; Richter 1979). The term complementizer agreement emerged more recently in the field of generative research (e.g. Bayer 1984; Fuß 2014; van Koppen 2017; Watanabe 2000; Zwart 1997). However, despite its name, the phenomenon is not restricted to complementizers in a narrow sense (i.e. subordinating conjunctions), but rather connected to the C-position in a hierarchical syntax tree (see Weiß 2005: 148–149). As Examples (1) to (5) from the Bavarian language area show, the agreement morpheme can attach to different classes of words or phrases at the left sentence bracket of embedded sentences, including: subordinating conjunctions (like ob ‘if’, weil ‘because’), embedded wh-questions (like warum ‘why’, wann ‘when’, wer ‘who’), wh-items as part of PPs (like bis wann ‘until when’, mit wem ‘with whom’), wh-items as determiners (like wie viel ‘how much/many’ or relative pronouns). While (1), (2), and (5) represent examples of “simple” word-size C-elements, (3) and (4) more “complex” C-elements in the form of a wh-DP and a wh-PP.¹

1. For further discussion of the complexity of C-Elements, see e.g. C-elements Bayer & Brandner (2008: 89) or Schallert et al. (2018: 28). In limiting ourselves to only syntactic complexity, we follow previous research into variation in the C-position. As this syntactic complexity is often reflected in phonological complexity, we do not want to rule out that these may also be factors for the occurrence of CA and DFC, yet do not pursue this here.

(1) Subordinating conjunction showing CA in 2PL (dialect example from Neckenmarkt, Burgenland, Austria, in the South-Central Bavarian transition area, DiÖ Corpus)

[...], ob-s es morgen arbeiten werd-ts
[...], if-2PL you tomorrow work will-2PL ‘if you will work tomorrow’

(2) Word-size wh-element functioning as a pro-form showing CA in 2PL (dialect example from Hall, Tyrol, Austria, in the South Bavarian area, Lenz et al. 2014: 12)
[...], warum-s es nit in Urlaub fohr-ts
[...], why-2PL you not in vacation go-2PL
‘why you don’t go on vacation’

(3) wh-PP including a simplex wh-word showing CA in 2sg (dialect example from Neckenmarkt, Burgenland, Austria, in the South-Central Bavarian transition area, DiÖ Corpus)
[...], bis wann-st morgen die Hausübung fertig ho-st
[...], until when-2sg tomorrow the homework done have-2sg
‘until when you will have finished your homework tomorrow’

(4) wh-DP including a full wh-phrase showing CA in 2sg (dialect example from Allentsteig, Niederösterreich, Austria, in the East Central Bavarian area, DiÖ Corpus)
[...], wie vü Freind-st nächstes Jōhr am Hof höfa wir-st
[...], how many friends-2sg next year at farm help will-2sg
‘how many friends will you help on the farm next year’

(5) Relative d-pronoun ‘whom’ showing CA in 2pl (dialect example from Gmunden, Upper Austria, Austria, in the East Central Bavarian area, Gruber 2008: 41)
De Frau, de-s (es) gseng hob-ts kenn i.
the woman whom-2pl (you) seen have-2pl know I
‘I know the woman, whom you have seen’

The parentheses in Example (5) indicate that the sentence is grammatical both with and without the full 2pl pronoun ‘es’ (or equivalent pronouns for the according combination of person and number). If the personal pronoun is not present in an utterance, it is unclear to what extent an interpretation, as an agreement morpheme, is appropriate. In such instances, an alternative interpretation would consider this morpheme a cliticized pronoun -es. Thus, without an additional disambiguating full pronoun (like in (1) and (2)), such cases constitute ambiguous instances of agreement (see also Chapter 3 and the discussion in Bohn & Weiß 2017). While such ambiguity can also occur with the 1pl subject pronoun/agreement morpheme ma/ ma (< mia/wir ‘we’) in Bavarian (see (6) versus (8)), instances of agreement with 2sg subjects are generally unambiguous (see (3) and (4) above), as they show CA-morpheme -st. Even if this -st-morpheme is

2. Glossing has been adapted in this and all following examples quoted.
3. CA also connects to the phenomenon of pro-drop, as it has been argued that CA-morphemes of pronominal origin license pro-drop. For a discussion of the interaction pf pro-drop and CA, see Weiß (2005: 154–157) and Weiß & Volodina (2018: 268–269).
interpreted as containing the cliticized pronoun -d/-t (\(<\) du/de ‘you’), its spirant- 
s tic s element cannot be explained through pronominal clitization.

(6) Ambiguous CA in 1PL (dialect example from Weißbriach, Carinthia, Austria, 
in the South Bavarian area, DiÖ Corpus)
[...], wie viel Freind dass-ma/=ma nächst Jahr im Urlaub treffen
[...], how many friends that-1PL/=we next year during vacation meet
wer-ma.
will-1PL
‘how many friends you will meet on vacation next year’

(7) Unambiguous CA in 2PL (dialect example from Tarrenz, Tyrol, Austria, in the 
Alemannic-Bavarian transition area, DiÖ Corpus)
[...], ob-ets es morgen arbeit-ets.
[...], if-2PL you tomorrow work-2PL
‘if you will work tomorrow’

(8) Unambiguous CA in the 1PL (dialect example from Oberwölz, Styria, Austria, 
South Bavarian area, DiÖ Corpus)
[...], wann-ma wir in Karl besuch-en.
[...], when-1PL we the Karl visit-1PL
‘when we will visit Karl’

Typologically, CA-morphemes can be classified depending on whether they are 
pronominal or verbal (Weiß 2005:154). Examples (3), (4), (6) and (7) show 
instances of such verbal agreement, whereby the agreement morphology is iden- 
tical between verb and C-position. In our data, this is generally the case with the 
2SG (see (3) and (4)). Regarding the 2PL, in Bavarian varieties, such a formal iden- 
tity is only documented for North Bavarian varieties that show an -ts-morpheme 
both for CA and verbal agreement (see Lenz et al. 2014: 13, below Figures 1 and 2).

Central Bavarian varieties, however, generally show CA with a mismatch between 
the verbal affix -ts and the CA-morpheme -s.\(^4\) This connects CA to the phenom- 

don of double agreement (DA), where verbs show different inflectional markers 
in dependence of their position (Zwart 1997; Weiß 2018). As we later discuss, DA 
may explain the historical emergence of CA. Returning to the forms of 2PL CA, 
(7) shows a previously undocumented variant of CA occurring in the Alemannic-

Bavarian transition area, which shows identical -ets morphemes for both CA and 
sentence final verbs. For the 1PL, (6) shows further morphological identity 
between verbal and (ambiguous) C-position-agreement. Yet (8) illustrates that

\(^4\) For a further example of mismatch in the 2PL between the verb ending -t and a CA-
morpheme -s see Lenz et al. (2014:13).
these morphemes may also differ for the 1pl: the CA-morpheme -ma (of pronominal origin) differs from the inflectional ending -en (for further examples of different inflectional morphemes for CA and verb inflection, see Weiß 2005:153, Bohn & Weiß 2016:162). But for the 1pl, there are also documented instances of CA where the agreement morpheme cannot be connected to pronouns but only mirror the verbal inflection (see e.g. Weiß 2005:154). Such an instance is given in (9) with a 1pl subject. CA thus appears as a very heterogeneous phenomenon.

(9) Unambiguous CA with 1pl without pronominal element (dialect example from East Franconian, adapted from Weise 1907:204, also cited in Weiß 2005:154)

\textit{waal-n mer graad besann sen-n}

because-1pl we just together are-1pl

‘because we are together right now’

Weiß (2005:149) points out that, while most of the syntactically focused research on CA is based on data from Dutch and Flemish (e.g. Haegeman 1992; Zwart 1997; van Koppen 2005) or from Upper German dialects (like the Bavarian dialects which we will focus on, e.g. Bayer 1984), the phenomenon appears in most Continental West Germanic dialects. Dialectological studies on CA in the German language area have often limited themselves to documenting the phenomenon in a given area by providing examples (see Table 1). Regarding the different linguistic items (words and phrases) that introduce the subordinate clauses, the accounts vary from individual examples to rather extensive lists. Table 1 gives an overview of dialect areas and C-elements documented in dialectological literature (without any claim to being exhaustive).

An overwhelming majority of examples in the literature appear to discuss CA with simple conjunctions (particularly \textit{ob ‘if’, dass ‘that’, and wenn ‘if, when’}) and \textit{wh}-words, but there is also evidence of CA with relative pronouns. Mention of more complex complementizers is rare but does occur (e.g. \textit{immer wenn ‘always when’}, see Gruber 2008:112). At the present stage of documentation, it is uncertain whether this reflects its overall limited occurrence, areal differences in the occurrence of CA with different C-elements, or whether such instances are merely undocumented.

Although there are studies documenting CA containing more complex complementizers, the general influence of complexity of the occurrence of CA is largely unexplored. Bayer, in contrast with earlier statements, suggests that CA morphemes attaching to \textit{wh}-phrases is “dispreferred or downright ungrammatical” (Bayer 2015:23). However, Example (4), taken from our corpus, suggests that occurrence with phrasal elements it is possible (see also Schallert et al. 2018:49–50). Further, there is uncertainty about whether there is an interdepen-
Table 1. Subordinating elements (words/phrases) showing CA described in previous research on German dialects

<table>
<thead>
<tr>
<th>Language area</th>
<th>Simple wh-word (question)</th>
<th>Relative elements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Western) Low German(^a)</td>
<td>dass ’that’</td>
<td>wann ’when’</td>
</tr>
<tr>
<td>(Western) Central German(^b)</td>
<td>als (wie) ’than’, dass ’that’, ob ’if’, weil ’because’, wenn ’if, when’</td>
<td>was ’what’, ’how’, wo ’where’</td>
</tr>
<tr>
<td>(Eastern) Central German(^c)</td>
<td>bis ’until’, ehe ’before’, dass ’that’, weil ’because’, wenn ’if, when’</td>
<td>wann ’when’, was ’what’, wie ’how’, wo ’where’</td>
</tr>
</tbody>
</table>

---

a. For Westfalian (Soest) see Holthausen 1886 (as cited in Weiß 2005:150); for Lower Franconian (Mülheim an der Ruhr) see Maurmann 1898:68 (as cited in Weiß 2005:150).
b. For Hesse (Lorsch) see Mottausch 2009:157; for Central Franconian see Newton 1990:174; for Ripuarian see München 1904:161 (as cited in Weiß 2005:150); for Wisconsin Heritage German (Rhine Franconian) see Bousquette (2014:565).
c. For Upper Saxon (Großpaschleben) see Richter (1979:532).
d. For East Franconian (Egerland, Upper Palatine) see Weise (1907:200); for Central and North Bavarian see Pfälz (1907:10); for Central Bavarian (Gmunden) see Gruber (2008:101–114); for East Central Bavarian (Landau) see Altmann (1984:200).

Bayer & Brander (2008:88–89) propose a hierarchy for the occurrence of DFC that ranks wh-words and phrases. This hierarchy assumes an increasing suitability for DFC from strictly (simplex) word size wh-elements (e.g. wer ’who’) over
(internally) complex *wh*-words involving e.g. a quantifier or dative case (e.g. *wie viel* ‘how many’, *wem* ‘who.DAT’) to *wh*-phrases like *wh*-PPs or *wh*-DPs (e.g. *bis wann* ‘until when’, *wie viele Leute* ‘how many people’ or *welche Farbe* ‘which color’). A study of acceptability judgements by Schallert et al. (2018) generally confirms this hierarchy of DFC-affinity among the various C-elements. However, they also suggest that there is strong regional variation in what is acceptable, and point to the relevance of the syntactic function (argument vs. adjunct) of the C-element in dependence of its complexity, as well as to animacy as a factor that promotes DFC. Weiß & Strobel (2018: 21) suggest that both DFC and CA (as well as other phenomena) are dependent on the permeability of the left periphery. They investigate data from Hesse and find an intraindividual interdependence between the two phenomena: While only a minority of the investigated speakers show DFC (10.92%) and CA (27.29%), more than half (50.42%) of the speakers who show CA also show DFC. Although DFC and CA have thus been sporadically brought into connection with each other (Schallert et al. 2018 49–59, Bayer 2015: 23), their immediate interaction has not been explored in depth and empirical studies that focus explicitly on this question are still a desideratum. In line with previous observations (Lenz et al. 2014: 12–13; Bayer 1984: 234), our data indicates that in instances where both DFC and CA appear, the agreement morpheme attaches to the second element (at the right border of the C-position), as seen in (11).

(10) DFC without CA with 2pl (dialect example from Neckenmarkt, Burgenland, Austria, in the South-Central Bavarian transition area, DiÖ Corpus)³

[...] *wie long ois moreng schlofen werd-ts.*
[...] how long C tomorrow sleep will-2pl
‘how long you will sleep tomorrow’

(11) DFC and unambiguous CA with 2pl (dialect example from Tarrenz, Tyrol, Austria, in the Alemannic-Bavarian transition area, DiÖ Corpus)

(DiÖ Corpus)

[...] *wiaviel Schulde dass-ets iberaht hob-ets.*
[...] how-much debt that-2pl in the first place have-2pl
‘how much you are in debt in the first place’

The central linguistic factors immediately regulating the occurrence of CA in the different languages and areas are the categories of person and number (see Weiß 2005: 150–151). According to Weiß (2005: 149–150), a minimal and most widely

5. The sentence further shows an instance of pro-drop. Based on further recordings from Neckenmarkt, *ois* should not be interpreted as a variant of the pronoun *es*, as speakers use *ois* in combination with e.g. 2sg and 1pl pronouns, too.
spread instance of complementizer agreement exists in the 2sg. This has recently been restated by Bohn & Weiß: “In den allermeisten deutschen Dialekten ist die Komplementiererflexion […] auf die 2. Pers. Sg. beschränkt und damit auf den Fall, in dem die Flexionsendung zumindest partiell pronominalen Ursprungs ist, d.h. durch Reanalyse des Subjektklitikums als (Teil der) Flexionsendung entstanden ist – das -t geht zurück auf eine Fehlsegmentierung des enklitischen thu”

However, there are deviations from this typological rule, e.g. in Hesse (Western Central German): As Bohn & Weiß (2017) show, the Hessian inflectional CA-ending appears more frequently in the 2pl than in the 2sg and thus provides a noteworthy contrast to the described pattern of the 2sg as the minimal form of complementizer agreement. Their analysis indicates that with the subordinating conjunction ob, in the 2pl a -d morpheme occurs in the entire area of Hesse, and also in areas where the use of the inflected variant appears more frequent than the uninflected variant. Additionally, they examined the occurrence of the agreement morpheme -st in the 2sg with wenn ‘if, when.’ Their 2sg data shows areas in central Hesse where their respondents did not indicate the use of agreement. Generally, the occurrence of agreement appears as a minority variant in the 2sg. It is reasonable to ask whether this CA paradigm pattern is particular to Hesse or whether it appears in other dialect areas as well. While the 2sg (despite the contradicting evidence from Hesse) thus shows a particular typological affinity for expressing CA, it is also worth mentioning that there are varieties where agreement is reported to extend over all persons (e.g. in Lapscheure dialect, see Haegeman 1992; Weiß 2005). Additional empirical data could present a more precise picture of the areal patterns of frequency of different forms of CA. They could also provide a better understanding of the emergence of CA in general, which is, as previously noted, a typologically rare phenomenon.

Much of the existing research on CA has focused on explanations as well as structural properties. Variationist questions, in contrast, have found limited consideration, and research with a broad empirical foundation is an exception. Lenz et al. (2014) conducted a first areal study of CA and the adverbal wh-expression warum ‘why’ with 2pl inflection for the Bavarian dialects (in Austria, Germany, and the Italian province South Tyrol). The study used questionnaires and received responses from 450 participants (347 younger aged 20–30, 103 older aged 65+) from 248 locations (see Figures 1 and 2). The results emphasize the importance of

---

6. In the overwhelming majority of German dialects, complementizer inflection is limited to the 2sg, and thus to the instance where the inflectional ending is at least in part of pronominal origin, i.e. where it emerged through reanalysis of the subject clitic as (part of the) inflectional ending – the -t goes back to a missegmentation of the enclitic thu. [translation M.F]
areality in the context of the Bavarian dialects and its three main subareas (Northern, Central and South Bavaria). First, there is a marked difference between the North and South in that CA occurs rarely in the South Bavarian area. Second, there are differences in the CA-morpheme for the 2PL, with the North Bavarian area showing a -ds/-ts, the Central Bavarian area showing only -s. Beyond regional patterns, the study also is first to address the intergenerational dynamics of CA. By contrasting the responses from the two age groups, it provides an insight into whether the phenomenon is stable across generations, or whether its use has changed. As Figures 1 and 2 indicate, the latter is not the case, and the phenomenon appears to have remained largely stable, regarding both its general regional occurrence as well as its morphological form.

Figure 1. Older participants’ (n=103) responses from a translation task targeting warum (‘why’) (adapted from Lenz et al. 2014: 28). warum\(^{\text{3s}}\) (‘why-3s’) = 19.4%, warum-\(d^{\text{3s}}\)s/-t\(^{\text{3s}}\)s/-z (‘why-3PL’) = 2.9%
Younger participants’ \( n = 347 \) responses from a translation task targeting \textit{warum} (‘why’) (adapted from Lenz et al. 2014: 29). \textit{warum}\(^s\) (‘why-2PL’) = 32.9%, \textit{warum-}\textit{d}'s/-t's/-z (‘why-2PL’) = 5.8%

The existing research, which focuses on the phenomenon in dialects, contains a further unanswered question: Is CA a phenomenon that is limited to dialects, or does it also occur in further linguistic repertoires. This seemingly trivial question is significant for understanding the spectrum of the “vertical” dimension of language between standard and non-standard registers. We address this and other questions in our following analysis.

2.2 CA: Explanations and emergence

Numerous theories attempting to explain CA have previously been proposed. Weise (1907: 201–205) refers to explanations that attribute contact with Slavonic languages, as a euphonic sound, as an distant associative effect of verb endings, as a product of enclisis, as well as a sheer phonetic process. Rinas (2006: 115)
provides a comprehensive overview of historical explanations for CA to date and refers to two further explanations from Dutch philology not previously mentioned: CA as a lapse in speech production that anticipates the verb ending, as well as a means of bridging the distance between sentence subject and finite verb. In contrast, Pfalz (1918), in accordance with the title of his article “Suffigierung der Personalpronomen im Donaubairischen” (‘suffixation of the personal pronouns in Danube Bavarian’) was likely the first to connect the phenomenon to clitic pronouns. However, he concedes that the ending -sd in 2SG cannot be explained by enclisis and interprets its occurrence as a transmission of the verb ending. The explanation through clitization holds currency to this day, as e.g. Lenz et al. (2014) also suggest that the regional occurrence of the complex plosive-spirans inflectional morpheme (-ts/-ds/-z) may be connected to the 2PL pronouns diats/deeds/diids (standard German ihr ‘you.2PL’). There have also been attempts to extend this explanation to the 2SG. Fuß (Fuß 2004) assumes a 2SG clitic pronoun -st, which is identical to the verb inflectional morpheme that was subject to reanalysis (see also Bayer 1984: 230). He supports this hypothesis using evidence from emergence of the new verb ending -st through reanalysis of the earlier -s ending and the pronoun tu in contexts where the pronoun follows the verb. Rinas (2005: 34–50) provides a very similar analysis that focuses on data from child language acquisition to provide a general account of reanalysis in the 2SG. He proposes that child language learners may interpret their input in such a way that they arrive at a system of 2SG pronouns that distinguishes an (stressed) du, occurring sentence initially (du kommst ‘you come’), and a non-stressed ste (komm ste ‘come you’), occurring sentence internally. This non-stressed pronoun then, in accordance with the general patterns of pronoun placement in German, can only appear sentence internally at the immediate beginning of the Mittelfeld.

For Rinas, this is the foundation for a reanalysis that may ultimately result in the emergence of CA. Weiβ (2005: 159–161) lays out a similar path to explain the origin of the 1PL inflectional morpheme -ma. This morpheme occurs as a CA-morpheme, as well as a morpheme of verbal inflection, and emerged from reanalysis of the homophonous subject clitic. His account includes a description of the different stages of grammaticalization that the reanalysis of this morpheme has reached in different varieties of Bavarian. These begin with the status of a mere subject clitic that cliticizes onto the complementizer or a verb in inverted contexts (12). In the second stage (see (13)), the clitic is in the process of reanalysis as it may be reduplicated with a full form in main clauses. Following this, (see (14)) -ma has the status of an inflectional element that is limited to the C-position but is obligatory, even in instances where the full pronoun is present. In the final stage (15), this new inflectional suffix also appears in sentence final positions. Both
the explanations for the 2sg and the 1pl are compatible and we can extend them to the 2pl by way of the pronoun es and its clitic form -s (Lenz et al. 2014:11–12). Notably, the path sketched for the 1pl involves an intermediate stage of DA with two different verb morphemes that appear dependent on the position of the finite verb in a clause (see also (8) above, Weiß 2018:135–136).

\[(12)\] \text{wuo=me han} \quad \text{where=1pl have} \quad \text{‘where we have’} \quad \text{(adapted from Weiß 2005:160)}

\[(13)\] \text{mia liŋ=ma} \quad \text{we lie=1pl} \quad \text{‘we lie’} \quad \text{(adapted from Weiß 2005:160)}

\[(14)\] \text{wa-ma mia hunt ha-nd} \quad \text{because-1pl we dogs are-1pl} \quad \text{‘because we are dogs’} \quad \text{(adapted from Weiß 2005:161)}

\[(15)\] \text{wa-ma bmaid ha-ma} \quad \text{because-1pl thought have-1pl} \quad \text{‘because we have thought’} \quad \text{(adapted from Weiß 2005:161)}

Building on this sequence, Weiß (2018:148–150) has proposed a more general theory of the emergence of CA that involves three stages. Precondition is the existence of clitic pronouns.\(^7\) In a second step, these clitics, as described above, may develop into inflectional forms that are limited to verbs in second position, i.e. DA. In a third and final step, child language learners may then associate the new inflectional ending with the C-position rather than only with verbs in this position, thus leading to the emergence of CA. It is only after the C-position has been associated with the properties of inflection, through such analogical extension, that further expansion of the paradigms can take place to show verbal CA, as e.g. in (9) above (Weiß 2018:150). The emergence of CA on this path may have been supported by the emergence of new pronominal forms through the interaction of pronouns and verbal inflection that is attested in many varieties of German (Weiß 2018:149). This explanation is general as much as it allows for the heterogeneity of CA phenomena, e.g. the different forms of 1pl CA in (9) and (14). A drawback of this theory is that it does not explain CA paradigms encountered in at least some varieties of Dutch, e.g. eastern dialects that only show 1pl CA with a morpheme -e that does not correspond to a clitic 1pl pronoun (see e.g. Hoekstra & Smits 1998:191–192). While the theory may therefore turn out not to be universally

\(^7\). For a discussion of the pronominal cycle and the development from full to clitic pronouns, see Weiß (2015).
applicable to all instances of CA in Germanic, it may still hold for certain regions. We therefore will apply it to the Upper German varieties spoken in Austria here. In doing so, we hope to contribute to understanding whether grammaticalization uniformly spreads from a single form of CA, e.g. the frequently documented 2sg, or whether CA can emerge independently from any clitic pronoun.

3. Analyses of complementizer agreement in Austria

The recent empirical studies of CA in German dialects (see Bohn & Weiß 2017 for Western Central German, Lenz et al. 2014 for Bavarian) address questions which were not addressed in earlier work. However, their reliance on indirect methods has inherent limitations. The present study extends beyond this because, as the first of its kind, it relies on a corpus of spoken language to explore the regional and inter- as well as intra-speaker variation, including data of different C-elements in different inflection contexts. The following section outlines the greater context of the research project that surrounds this study on CA. We then discuss the data for CA elicited in two survey settings of the project. The focus is on the results from completion tasks in a “Language Production Experiment” (LPE) that identifies linguistic, situational, and areal factors. In addition, we discuss first results from the investigation of open conversations from a subsample of the investigated locations. We then discuss whether this data fits a theory of the emergence of CA.

3.1 Larger project context

As mentioned above, this study uses a corpus that is currently being created in the context of the SFB “German in Austria. Variation – Contact – Perception” (for a detailed discussion, see Lenz 2018). Project part 03 (“Speech repertoires and varietal spectra” with a focus on rural areas of Austria) records speakers in up to seven different settings that target different sections of speaker’s individual linguistic repertoires. Two of these settings are conversational: first, a (rather formal) interview with a linguistic field worker, and second, a “conversation among friends” between two or more speakers in absence of an interviewer. The remaining five settings are more controlled elicitation methods that target a range of specific linguistic phenomena and consist of two runs of LPEs containing different task types targeting morphological and syntactic phenomena, two runs of translation tasks and last, two “reading out loud” tasks. In the following section, we rely

8. The programme is supported by the Austrian Science Foundation (FWF), with the first funding period from 2016 to 2019. See the project homepage: www.dioe.at.
on data from language production experiments (LPEs) and conversations among speakers from a given research location recorded in absence of a field worker.

As Figure 3 shows, the dialects of Austria are classified into two larger groups: Bavarian and Alemannic. The largest part of Austria is covered by dialects classified as Bavarian. These are subdivided into a Central and Southern subgroup, whereas the Northern Bavarian area is located in Germany. The area between these two regions is considered a transition area. The Alemannic varieties are only spoken in the far West of Austria (but cover German speaking Switzerland and South Western Germany), and connect to the Bavarian area through a further transition area. The data used in our analysis comes from speakers from 13 rural locations surveyed by our project part (PP03) that spread across Austria and represent the different dialect areas and their subregions (see Figure 3). Larger areas are represented by multiple locations. These locations are rural villages with about 500 to 2000 inhabitants that are not in close vicinity of urban centers as Vienna, Graz, and others.

Figure 3. Major Austrian dialect areas (following Wiesinger 1983) and research locations of the SFB project part 03

In each location, the project part aimed at a core group of 10 speakers. The central criterion for these speakers is their “autochthony”. This requires that speakers grew up in the location, that if they no longer live in the location, identify with it, and have continuous ties to friends, family etc. in the location, as well as at least one parent who is native to the location or its immediate vicinity. Of the 10 participants, 8 are recruited from a younger (18–35) and 2 from an older (60+) age group. While the older participants represent classical NORMs
and NORFs and have a lower degree of formal education, the younger speakers generally show a greater regional mobility and are diverse in their educational background. Recruitment of participants also considers speakers’ level of formal education. Because the Austrian education system is rather complex, this distinction is based on whether participants may attend an institution of higher education (university) or not. Only for the younger generation, speakers with a higher level of formal education are recruited after and make up half of the participants. All groups should contain equal parts of male and female participants. While this group of 10 speakers fails to be representative in a statistical sense, they provide a cross section of speaker groups from a given location that focuses on younger speakers in particular. Despite all efforts, we could not recruit the targeted speakers in all locations. Particularly for the younger generation, recruitment efforts met considerable sociodemographic variation. For example, it was not possible to find two young female speakers in all locations with a lower level of formal education that were willing to participate in our (extensive) surveys. To reach the target goal of 10 speakers, the project interviewed speakers of the other sex or with a different educational background.

3.2 Data elicitation: Language production experiments (LPEs)

The analyses of CA we present primarily rely on data collected using LPEs (see Lenz et al. 2019). These experiments use auditory and audiovisual stimuli that are presented with the support of computers and aim at the production of spoken language. The participants’ responses are recorded and subsequently transcribed and annotated for further analysis. In comparison with oral questionnaires administered by field workers, the LPE’s rely on computers to present the stimuli which allows for a higher degree of inter- and intra-individual repeatability (see Breuer & Bülow, 2019). In the context of data collection, the experiments on CA are embedded into further experiments that target different syntactic and morphological phenomena and appear in a pseudo-randomized order. To investigate a speaker’s linguistic repertoires on the dialect-standard axis, each experiment is administered targeting both the speaker’s standard German and their dialect. While the stimuli used in the LPEs targeting the standard register consist of recordings of a trained newscaster from the Austrian Broadcasting Corporation (ORF), the stimuli targeting the dialect are recordings from speakers with proficiency in the dialect of the location in question. The experiment design generally builds on previous research and targets specific factors that are known or hypothesized to influence a phenomenon (see Lenz et al. 2019; Lenz 2009, Lenz 2016). Regarding CA, the first investigated factor concerns the type and complexity of the element in C-position. We assume that elements which are more structurally simple show CA more frequently
than complex wh-phrases (see above). In the experiments analyzed here, we target the production of clauses introduced by

- the subordinating conjunction ob 'if'
- the simple wh-adverb wann 'when'
- the complex wh-element wie viel as a determiner (wie viele + NP 'how many + NP')

Further factors considered in the experiments are the verbal categories person and number. Previous research indicates that clauses with 2sg and 2pl show CA in particular, while it is less frequent with 1pl. Accordingly, the experiments target these three inflectional contexts. As a third factor, previous research indicates regional differences in the occurrence of CA. Within Austria, Central Bavarian varieties are particularly likely to show CA, while we expect it to a lesser extent in South Bavarian and Alemannic varieties. The factor of areal variation is addressed through selection of participants from according locations. The data we discuss stems from 13 locations for the experiments targeting local dialects and 5 locations for those targeting standard language (see Table 2 and 3). For every location, there are between 9 and 13 participants. For the dialect experiments, we consider data from 144 speakers (106 (73.6%) younger, 38 (26.4%) older; see Table 2), for those targeting standard German 58 speakers (41 (70%) younger, 17 (30%) older; see Table 3) who also participated in the dialect experiment. The inclusion of a higher number of younger than older speakers is motivated by considerations from the larger project context and tries to capture a broader range of sociodemographic variation with younger speakers.

Table 3. Speakers per location and generation in the LPE targeting Standard German

<table>
<thead>
<tr>
<th>Dialect area</th>
<th>Location</th>
<th>Speaker group</th>
<th>Total speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Bavarian</td>
<td>Neumarkt an der Ybbs</td>
<td>Older 3</td>
<td>Younger 8</td>
</tr>
<tr>
<td>South-Central Bavarian Transition Area</td>
<td>Neckenmarkt</td>
<td>Older 5</td>
<td>Younger 7</td>
</tr>
<tr>
<td>South Bavarian</td>
<td>Weißbriach</td>
<td>Older 2</td>
<td>Younger 9</td>
</tr>
<tr>
<td>Alemannic-Bavarian Transition Area</td>
<td>Tarrenz</td>
<td>Older 4</td>
<td>Younger 8</td>
</tr>
<tr>
<td>Alemannic</td>
<td>Raggal</td>
<td>Older 3</td>
<td>Younger 9</td>
</tr>
<tr>
<td>Total speakers</td>
<td></td>
<td>Older 17</td>
<td>Younger 41</td>
</tr>
<tr>
<td>Dialect area</td>
<td>Location</td>
<td>Speaker group</td>
<td>Total speakers</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Central Bavarian</td>
<td>Allentsteig</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Gaweinstal</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Taufkirchen an der Pram</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Neumarkt an der Ybbs</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Steyrling</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>South-Central Bavarian Transition Area</td>
<td>Neckenmarkt</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Passail</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Hüttschlag</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>South Bavarian</td>
<td>Oberwölz</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Weißbriach</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Tux</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Alemannic-Bavarian Transition Area</td>
<td>Tarrenz</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Alemannic</td>
<td>Raggal</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Total speakers</td>
<td></td>
<td>38</td>
<td>106</td>
</tr>
</tbody>
</table>

Our CA experiments aim to elicit subordinate clauses in three sets of three tasks each for a total of nine stimuli: Each experiment stimulus targets the subordinate clause introductions discussed above (ob ‘if’, wann ‘when’, wie viele + NP ‘how many + NP’) in combination with a verbal inflection (2SG, 1PL, 2PL).

The LPEs use a combined auditory and visual stimulus. The visual stimulus consists of images containing an embedded written context that correspond to a series of events narrated by a recorded voice in the auditory stimulus component (see Figure 4). The first two events provide a context of past and present events (e.g. *Letztes Jahr habt ihr 3 Leuten auf dem Hof geholfen. Dieses Jahr habt ihr 5 Leuten auf dem Hof geholfen. ‘Last year you helped 3 people on the farm. This year you helped 5 people on the farm.’). The third event is placed in the future by the text in the visual stimulus (e.g. *nächstes Jahr ‘next year’) and bears a question mark in the visual stimulus. After uttering *Ich frage mich, …* (‘I wonder, ...’), the narrator stops and thus leaves completion of the sentence to the participant, who may e.g. respond *wie vielen Leuten ihr nächstes Jahr auf dem Hof helft* (‘how many people you will help on the farm next year’). For our analysis of CA, responses are considered “relevant” if the speakers complete the given matrix clause using the target C-element (ob ‘if’, wann ‘when’, wie viele + NP ‘how many + NP’) as introduction of a subordinate sentential construction, i.e. an embedded sentence with the inflected verb in final position. The tasks eliciting standard
and dialect registers closely resemble each other (for a detailed description, see Lenz et al. 2019). They are, however, not identical. In the context of the survey, the experiments are divided by a set of translation tasks that take approximately 30 minutes and thus do not immediately follow each other. We exclude responses with the inflected verb in first or second position from the analysis because (in line with both previous research and our data) they do not constitute an environment that permits the occurrence of CA.

Figure 4. Example of a visual stimulus in the LPE for CA. Parallel to the visual stimulus an auditory stimulus is given (Letztes Jahr habt ihr 3 Leuten auf dem Hof geholfen. Dieses Jahr habt ihr 5 Leuten auf dem Hof geholfen. ‘Last year you helped 3 people on the farm. This year you helped 5 people on the farm.’) that cuts off after the clause Ich frage mich, ... (‘I wonder, …’) and thus lets the participant complete the sentence

3.3 Analyses of experimental data: CA-variation in Austria

In the following section, we discuss the data elicited through the LPEs. While a discussion of elicitation success is part of the analysis, we want to preface the discussion with some practical considerations. As previously mentioned, CA-morphemes show areal variation. In our initial analysis, we do not discuss the differences in their morphological form, but only distinguish whether responses do or do not show CA, or whether there is an ambiguous instance where it is impossible to distinguish CA from cliticized pronouns (see Section 2).

While the setup of our LPEs is repeatable and aims to elicit one answer per prompt per participant, speaker’s responses to the stimuli vary. While some speakers respond to the prompt as intended by the design, other speakers provide
multiple responses, e.g., by repeating their answer using a different verb.\textsuperscript{9} Multiple responses may also indicate both the presence and absence of an investigated phenomenon in a speaker’s linguistic repertoire.\textsuperscript{10} In such instances, we counted both responses offered by the participant, assuming that they represent equally acceptable alternatives. However, some participants responded multiple times to a single stimulus or used the same C-elements and personal forms to respond to different stimuli. To avoid overrepresentation of such respondents, we did not count multiple responses that offered identical instances of agreement, non-agreement and ambiguous instances with the same C-element by the same participant multiple times. E.g., if a participant would give two responses to a stimulus that both showed agreement with the same C-element, the following analysis counts this as a single instance of agreement.\textsuperscript{11} While a single speaker may thus provide responses for more than one of the possible outcomes, no single speaker provides multiple responses towards the same outcome. Therefore, instances of (ambiguous/non-) agreement for a given C-element indicate the number of speakers that used an according form in at least one of their responses. We begin our analysis with a discussion of linguistic and contextual factors influencing CA. Following this, we move to a closer investigation of areality, which leads us to a discussion of how the data elicited through the LPE relate to data recorded in conversations.

\subsection*{3.3.1 Linguistic factors}

We begin the discussion of our findings with a focus on linguistic factors influencing the variation of CA within the dialects of Austria. Figure 5 gives a first overview of the LPE-D responses from all 144 speakers regardless of their region or age. The chart shows the total responses for different subordinate clause introducing elements with and without DFC (see Chapter 2) and distinguishes CA

\textsuperscript{9} A participant from Allentstein (Lower Austria, East Central Bavarian) e.g. gave a first response “I frog mi, ob-s/=s murgen hackelts” (‘I wonder if-2PL/=you will work tomorrow’) that they then corrected to “oder i frog mi, ob-s/=s murgen wos arbeits” (‘I wonder if-2PL/=you will work something tomorrow’), slightly altering the construction and using a different word for work (arbeiten vs. hackeln).

\textsuperscript{10} A participant from Allentstein (Lower Austria, East Central Bavarian) e.g. showed ambiguous CA in one response using wann: “I frog mi, wann-/=s in Michael höfen werdts” (‘I wonder when-2PL/=you will help Michael tomorrow’). To a different stimulus, their response showed an unambiguous instance of CA: “I frog mi, wann-s es in Korf besuchen werdts.” (‘I wonder when-2PL you will visit Karl’).

\textsuperscript{11} E.g a participant from Oberwölz (Styria, South Bavarian) responses to two different stimuli both showed 2PL CA with ob: “ob-s ihr morgen wos arbeits werdts” (‘ob-2PL you will work something tomorrow’) and “ob-s ihr morgen in die stodt fohrts” (‘if-2PL you will drive to the city tomorrow’). These two responses only count as a single instance of CA with ob.
from the alternative variants of ambiguous and non-CA. The dialect data shows that DFC only occurs with the simple wh-element wann ‘when’ \((n=13)\) and, more frequently, with more complex wh-phrases including wie viele + NP ‘how many + NP’ \((n=69)\). With ob ‘if’, there is no evidence of two elements in C-position. In those instances where DFC appears, the second C-element is generally either dass ‘that’ or als ‘when’/‘than’ (dialectally pronounced as ois). Regarding CA, unambiguous instances of CA do not appear with DFC and are limited to instances where there is only one C-element, while instances of ambiguous and non-agreement occur with all C-elements.\(^\text{12}\)

![Figure 5. Absolute frequencies of CA and DFC in the LPE-D \((n=798)\) in all 13 locations, 144 speakers](image)

**Table 4.** Elicitation success rates in the LPE-D in all 13 locations, 144 speakers

<table>
<thead>
<tr>
<th>C-Element</th>
<th>Success rate</th>
<th>(432 possible responses assumed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ob ‘if’</td>
<td>344</td>
<td>79.6%</td>
</tr>
<tr>
<td>wann ‘when’</td>
<td>239</td>
<td>55.3%</td>
</tr>
<tr>
<td>wie viele + NP</td>
<td>213</td>
<td>49.3%</td>
</tr>
</tbody>
</table>

As we discuss in further detail below, the structure of the C-element is a strong factor in the occurrence of CA in the Austrian language area. In our following analysis, we set aside the interaction of CA and DFC and focus on the discussion

\(^{12}\) However, further data from our translation tasks (from standard into dialect) shows sporadic instances of unambiguous CA appearing with DFC in the Bavarian area (see also Bayer 1984: 235, Lenz et al. 2014: 12–13).
of CA with non-DFC. The comparison of the three different C-elements already allows for first observations on elicitation success. While there are 344 relevant responses including *ob* ‘if’, the numbers for *wann* ‘when’ (*n*=238) and *wie viele + NP* ‘how many + NP’ (*n*=209) are lower. Assuming one possible answer per participant (*n*=144) for 3 stimuli per C-element (144×3=432), this leads to success rates of 79.6% for *ob* ‘if’, 55.1% for *wann* ‘when’ and 48.4% for *wie viele + NP* ‘how many + NP’ (see Table 4). The difference in the success rates between the C-elements has multiple causes and reflects the open nature of the experimental approach that, while targeting a specific form, gives the participant a degree of liberty in his response. Here, the syntactic properties of the targeted C-elements influence the responses. While German *ob* ‘if’ only functions as a subordinate clause introduction, *wann* ‘when’ and *wie viele* ‘how many’ can introduce *wh*-questions with verbs in second position (e.g. *(Ich frage mich:) Wie vielen Leuten hilft ihr nächstes Jahr auf dem Hof?* ‘(I wonder:) How many people will you help on the farm next year?’ instead of *(Ich frage mich,) wie vielen Leuten ihr nächstes Jahr auf dem Hof hilft?’*(‘I wonder,) how many people you will help on the farm next year?’). Such verb second sentences cannot show CA and we therefore have to exclude them in our analysis. Further, the stimuli targeting *wann* ‘when’ and *wie viele* ‘how many’ elicited a greater breadth of responses than others. Participants e.g. responded with a more complex C-phrase *um wie viel Uhr* ‘at what time’ or with the less complex *wie viele* ‘how many’ (without an NP). While we again exclude these “unintended” responses here, they are not irrelevant in the narrow sense, but allow for the investigation of further C-elements. The same goes for responses that have other verbal inflections, e.g. 1sg. Even with this selected focus only on the responses aimed at through the elicitation, the data provides a sizeable corpus for the investigation of CA and DFC, in particular when considering the relative scarcity of these C-elements in open conversations (which we will demonstrate below).

Figure 6 focuses on the responses without DFC and distinguishes responses for explicit CA (“CA”), ambiguous instances (“ambiguous”), and non-CA (“no CA”) in embedded sentences with 2sg, 1pl, and 2pl. This provides a more nuanced pattern of CA. In the 1pl, ambiguous instances of CA are abundant, occurring with all three C-elements, but there is only a single unambiguous instance with *wann* ‘when’ (see (8) above). In contrast, there are no ambiguous responses in the 2sg.13 While agreement appears frequently with *ob* ‘if’ and *wann*

---

13. There are documented instances where the 2sg may exhibit -t morphemes following the C-position from Zurich German (i.e. Alemannic, see Weiß 2005:157). While Weiß proposes an interpretation as CA for Zurich German, Zwart (1993:257) does interpret the same data as non-CA. Applying our approach to ambiguity from the 2pl, we would classify such -t-morphemes
‘when’, there are only two instances of CA with *wie viele + NP* ‘how many + NP’ (see e.g. (4) above). For the 2*pl*, CA as well as non-CA and ambiguous instances appear. While both ambiguous and unambiguous instances of CA appear frequent with *ob* ‘if’ and *wann* ‘when’, they are again an exception with *wie viele + NP* ‘how many + NP’.

![Figure 6. Absolute frequencies of CA in the LPE-D (n = 727) in all 13 locations, 144 speakers](image)

Figure 7 represents the relative frequencies of the same data. The difference between 1*pl* and 2*sg* and 2*pl* remains drastic. As already stated, there is a near complete absence of unambiguous agreement with all three C-elements with the 1*pl*, occurring only in a single instance (1.2%) with *wann* (‘when’). In contrast to this, over half (55.7%) of the 1*pl* responses with *wie viele + NP* ‘how many + NP’ show ambiguous CA. This again stands against the responses from 2*sg* and 2*pl*. In the 2*sg*, unambiguous instances of CA account for only a 4.4% of the responses with *wie viele* ‘how many’, and only 10.8% of the responses with 2*pl* show CA or are ambiguous. 2*sg* and 2*pl* thus appear much more similar. Because of the ambiguous instances in the 2*pl*, a comparison of the results requires caution. Considering only the unambiguous instances, with *ob* ‘if’, CA appears more frequently in the 2*pl* (59.1%) than in the 2*sg* (42.6%). However, with *wann* ‘when’, this difference is leveled: with 48.6% CA with 2*sg* and 50% with 2*pl*, the frequencies are essentially equal. Only when we consider the ambiguous instances as instances of CA does the 2*pl* emerge as more frequent with both *ob* ‘if’ (71.8%) as ambiguous without further evidence of a co-occurring pronoun. However, our data does not show any such ambiguous morphemes for the 2*sg* with the three C-elements in question. Only in instances of DFC, where the second C-element ends on an /s/, i.e. with *dass* ‘that’ and *ois* ‘than’, does ambiguity arise as a the morpheme boundaries blur.
and wann ‘when’ (72.2%). However, without further research on the according varieties we refrain from doing so.\textsuperscript{14}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Relative frequencies of CA in the LPE-D ($n=727$) in all 13 locations, 144 speakers}
\end{figure}

To explain the differences between these findings, it is worth looking back at the grammaticalization sequence for the 1\textsubscript{pl} proposed by Weiß (2005, see (12) to (15) above). As Weiß (2005) notes, the non-emphatic 1\textsubscript{pl} personal pronoun ma has undergone grammaticalization towards a CA-morpheme only in some varieties of Bavarian. The abundance of ambiguous instances speaks to this status. While the ambiguity remains, the difference becomes apparent when contrasted with the 2\textsubscript{pl}, where ambiguous instances can also occur. Here, the CA morpheme -s, may be interpreted as a clitic 2\textsubscript{pl} pronoun -s (< Bavarian es ‘you’) in absence of second element that is a pronoun beyond doubt. These ambiguous instances are less frequent than with the 1\textsubscript{pl} not only with ob ‘if’ and wann ‘when’, but also with wie viele + NP ‘how many + NP’. The near complete absence of unambiguous CA with the 1\textsubscript{pl} across C-elements, combined with the appearance of the ma-morpheme (1\textsubscript{pl}) with wie viele + NP ‘how many + NP’, suggests that in the majority of Austrian varieties, we are not dealing with a CA-morpheme, but with a non-stressed pronoun ma (< Bavarian mia ‘we’). This indicates that in Austria, CA is more widely spread with 2SG and 2PL than with 1PL. Simultaneously,

\textsuperscript{14} Disappearance of the pronoun es in favor of the Standard German variant ihr could e.g. be considered a reason against an interpretation of -s as a clitic pronoun and instead for its interpretation as a CA morpheme. We will touch upon this briefly later. However, a detailed investigation of the dynamics of the pronominal system in the according varieties would go far beyond the scope of this contribution.
it contributes to confirming a second assumption: CA occurs less frequently with complex wh-phrases like *wie viele + NP* ‘how many + NP’ than with simpler C-elements like *ob* ‘if’ and *wann* ‘when’. We return to these findings as we discuss the data in more detail.

### 3.3.2 Situational factors

At this point, we want to briefly consider how situational factors influence the occurrence of CA. Figure 8 shows the responses from the LPE targeting participants’ intended Standard German (LPE-S) from five locations taken from the major Austrian dialect areas (Neckenmarkt, Neumarkt, Raggal, Tarrenz, Weißbrach; see Figure 3). As Table 5 shows, the tasks targeting *ob* ‘if’ are again the most successful and provide relevant responses to 144 (82.8%) of the 174 stimuli. In contrast, only 71 (40.8%) of the tasks aiming at *wann* ‘when’ and 76 (43.7%) of the tasks targeting *wie viele + NP* (‘how many + NP’) elicit relevant responses. This response pattern is generally comparable to the behavior observed with the LPE-D. None of the responses involved DFC, and as Figure 8 shows, CA is absent regardless of the C-element. This indicates that both CA and DFC are very much phenomena of non-standard varieties. Therefore, in the following section, we focus on data only from the experiments targeting speakers’ dialects.

![Figure 8](image)

**Figure 8.** Absolute frequencies of CA in the LPE-S (*n* = 291) in 5 locations, 58 speakers

### 3.3.3 Areal factors

After the initial discussion of general linguistic factors, we turn our focus to areal distribution of CA in the Austrian dialects and discuss the patterns of the three inflection forms: 1PL, 2PL, 2SG. First, we combine the results for all three C-elements, but again exclude instances of DFC. As the absolute numbers indicate, there is some variation in between locations in the number of relevant responses.
Table 5. Elicitation success rates in the LPE-S in 5 locations, 58 speakers

<table>
<thead>
<tr>
<th>C-Element</th>
<th>Responses</th>
<th>Success rate (174 possible responses assumed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ob ‘if’</td>
<td>144</td>
<td>82.8%</td>
</tr>
<tr>
<td>wann ‘when’</td>
<td>71</td>
<td>40.8%</td>
</tr>
<tr>
<td>wie viele + NP</td>
<td>76</td>
<td>43.7%</td>
</tr>
</tbody>
</table>

We see three main reasons for this: The slight variation in participant numbers for the individual locations (see Table 2), areal differences in the occurrence of DFC (and therefore exclusion of relevant responses from this analysis focused on CA without DFC), and a general fluctuation in elicitation success (or “creativity” of our speakers in responding to the stimuli). Therefore, the relative frequencies of CA may be influenced by these factors. This is e.g. the case in Tux where, more than in other locations, wann ‘when’ occurs with a second C-element. Our interpretation of the relative numbers therefore considers this and does not overly stress small differences in relative frequencies. Despite this, the data gives a very good account of the areal distribution of the different forms of CA.

Figure 9. Relative frequencies of 2sg CA in the LPE-D with C-elements ob ‘if’, wann ‘when’, and wie viele + NP ‘how many + NP’. Data from 13 locations, 144 speakers. For details, see Table 6

We begin with a discussion of the 2sg. Table 6 shows the combined occurrences of CA for all three C-elements in the 13 research locations. This data is the foundation for the charts in Figure 9. The map shows a marked regional pattern:
2sg CA is absent in the westernmost locations Raggal, Tarrenz, and Tux. While the South Bavarian locations Weißbriach (3.7%) and Oberwölz (18.2%) do show CA, they do so less frequently than even the location with the smallest percentage of CA in the Central Bavarian or South-Central Bavarian transition area, Gaweinstal (47.1%). Considering that the occurrence of only a single response with CA in Weißbriach, it seems possible that this is not a traditional feature of the Weißbriach dialect. While the according speaker is autochthonous to Weißbriach by our criteria, he has been living in Vienna for several years at the time of recording. The occurrence of CA with this speaker may thus indicate influence from other

### Table 6. Relative and absolute frequencies of 2sg CA in the LPE-D with C-elements ob 'if', wann 'when', and wie viele + NP 'how many + NP'. Data from all 13 locations, 144 speakers, $n = 234$

<table>
<thead>
<tr>
<th>Dialect area</th>
<th>Location</th>
<th>CA</th>
<th>ambiguous</th>
<th>no CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Bavarian</td>
<td>Allentsteig</td>
<td>17 (68%)</td>
<td>0 (0%)</td>
<td>8 (32%)</td>
</tr>
<tr>
<td></td>
<td>Gaweinsteal</td>
<td>8 (47.1%)</td>
<td>0 (0%)</td>
<td>9 (52.9%)</td>
</tr>
<tr>
<td></td>
<td>Taufkirchen an der Pram</td>
<td>16 (94.1%)</td>
<td>0 (0%)</td>
<td>1 (5.9%)</td>
</tr>
<tr>
<td></td>
<td>Neumarkt an der Ybbs</td>
<td>9 (50%)</td>
<td>0 (0%)</td>
<td>9 (50%)</td>
</tr>
<tr>
<td></td>
<td>Steyrting</td>
<td>9 (64.3%)</td>
<td>0 (0%)</td>
<td>5 (35.7%)</td>
</tr>
<tr>
<td>South-Central Bavarian Transition Area</td>
<td>Neckenmarkt</td>
<td>12 (63.2%)</td>
<td>0 (0%)</td>
<td>7 (36.8%)</td>
</tr>
<tr>
<td></td>
<td>Passail</td>
<td>7 (46.7%)</td>
<td>0 (0%)</td>
<td>8 (53.3%)</td>
</tr>
<tr>
<td></td>
<td>Hüttschlag</td>
<td>5 (50%)</td>
<td>0 (0%)</td>
<td>5 (50%)</td>
</tr>
<tr>
<td>South Bavarian</td>
<td>Oberwölz</td>
<td>4 (18.2%)</td>
<td>0 (0%)</td>
<td>18 (81.8%)</td>
</tr>
<tr>
<td></td>
<td>Weißbriach</td>
<td>1 (3.7%)</td>
<td>0 (0%)</td>
<td>26 (96.3%)</td>
</tr>
<tr>
<td></td>
<td>Tux</td>
<td>0 (100%)</td>
<td>0 (0%)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>Alemannic-Bavarian Transition Area</td>
<td>Tarrenz</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>23 (100%)</td>
</tr>
<tr>
<td>Alemannic</td>
<td>Raggal</td>
<td>0 (100%)</td>
<td>0 (0%)</td>
<td>15 (100%)</td>
</tr>
</tbody>
</table>

2sg CA is absent in the westernmost locations Raggal, Tarrenz, and Tux. While the South Bavarian locations Weißbriach (3.7%) and Oberwölz (18.2%) do show CA, they do so less frequently than even the location with the smallest percentage of CA in the Central Bavarian or South-Central Bavarian transition area, Gaweinstal (47.1%). Considering that the occurrence of only a single response with CA in Weißbriach, it seems possible that this is not a traditional feature of the Weißbriach dialect. While the according speaker is autochthonous to Weißbriach by our criteria, he has been living in Vienna for several years at the time of recording. The occurrence of CA with this speaker may thus indicate influence from other
Austrian varieties of German. Within the Central Bavarian and the South-Central Bavarian transition area, there is some fluctuation in the frequency. Against the occurrences in about every second response in Gaweinstal (47.1%) stands the almost universal occurrence of CA in Taufkirchen (94.1%). The remaining locations in these two areas range in between these and do not give a clear image of areal spread. Because response success with the three C-elements, (non-)occurrence of DFC and intergenerational dynamics may distort these results, we do not want to give the exact frequencies too much weight. Nonetheless, the general regional pattern appears robust. It is worth noting that 2sg CA by and large occurs along the lines of dialect areas (which rely on phonological criteria).

As previously discussed, the data for the 1pl shows an abundance of ambiguous instances of CA. Nonetheless, the investigation of areal distribution, shown in Figure 10 and Table 7, allows for some insights. As the map shows, ambiguous instances occur in all 13 locations investigated. While they generally account for over half of the instances, one location stands out from this pattern. This location is Raggal, in the Alemannic area, where only 10% of all instances are ambiguous. This differs from its neighboring location Tarrenz, located in the Alemannic-Bavarian transition area, where ambiguous instances make up 69.2% of responses. We find a likely explanation for this in an underlying difference in the Bavarian and Alemannic pronominal systems that momentarily leads away from the phenomenon of CA, but nonetheless is worthy of discussion. For
Table 7. Relative and absolute frequencies of 1\textsubscript{PL} CA in the LPE-D with C-elements *ob* ‘if’, *wann* ‘when’, and *wie viele* + NP ‘how many + NP’. Data from 13 locations, 144 speakers, *n* = 265

<table>
<thead>
<tr>
<th>Dialect area</th>
<th>Location</th>
<th>CA</th>
<th>ambiguous</th>
<th>no CA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Bavarian</td>
<td>Allentsteig</td>
<td>0 (0%)</td>
<td>21 (77.8%)</td>
<td>6 (22.2%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gaweinstal</td>
<td>0 (0%)</td>
<td>16 (80%)</td>
<td>4 (20%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Taufkirchen an der</td>
<td>0 (0%)</td>
<td>17 (89.5%)</td>
<td>2 (10.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pram</td>
<td>0 (0%)</td>
<td>16 (89.5%)</td>
<td>2 (10.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neumarkt an der</td>
<td>0 (0%)</td>
<td>16 (88.7%)</td>
<td>2 (13.3%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ybbs</td>
<td>0 (0%)</td>
<td>16 (89.5%)</td>
<td>2 (10.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Steyrling</td>
<td>0 (0%)</td>
<td>13 (86.7%)</td>
<td>2 (13.3%)</td>
<td></td>
</tr>
<tr>
<td>South-Central Bavarian Transition</td>
<td>Neckenmarkt</td>
<td>0 (0%)</td>
<td>16 (84.2%)</td>
<td>3 (15.8%)</td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>Passail</td>
<td>0 (0%)</td>
<td>14 (87.5%)</td>
<td>2 (12.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hüttschlag</td>
<td>0 (0%)</td>
<td>13 (76.5%)</td>
<td>4 (23.5%)</td>
<td></td>
</tr>
<tr>
<td>South Bavarian</td>
<td>Oberwölz</td>
<td>1</td>
<td>14 (53.8%)</td>
<td>11 (46.2%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weissbrach</td>
<td>0 (0%)</td>
<td>17 (76.5%)</td>
<td>6 (23.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tux</td>
<td>0 (0%)</td>
<td>10 (66.7%)</td>
<td>5 (33.3%)</td>
<td></td>
</tr>
<tr>
<td>Alemannic-Bavarian Transition Area</td>
<td>Tarrenz</td>
<td>0 (0%)</td>
<td>18 (69.2%)</td>
<td>8 (30.8%)</td>
<td></td>
</tr>
<tr>
<td>Alemannic</td>
<td>Raggal</td>
<td>0 (0%)</td>
<td>2 (10%)</td>
<td>18 (90%)</td>
<td></td>
</tr>
</tbody>
</table>

Bernese German (an Alemannic variety in Switzerland), Nübling (1992: 252) distinguishes a 1\textsubscript{PL} full pronoun *mir* ‘we’ and the clitic *mer* ‘we’. Jutz’ (1925: 272–273) description of the pronominal system of Southern Vorarlberg shows that there is a general variation in the 1\textsubscript{PL} pronouns in Alemannic speaking Austria but makes a similar system likely for the variety of Raggal. Jutz (1925) states that the 1\textsubscript{PL} pronouns are generally identical to the forms of the dative singular. These are *miər* ‘we’ in stressed positions and *mr* ‘we’ in unstressed position. Regarding the unstressed pronouns, some Alemannic varieties further show reduced or full vowels. His discussion of variation between different pronominal systems further
finds that stressed as well as unstressed 1pl pronominial forms frequently have an initial m-. Some Walser varieties (in Vorarlberg) have an initial w- in both forms. Additionally, some varieties in Southern Vorarlberg have forms ber ‘we’ or per ‘we’ with an initial plosive. Despite this considerable variation, Jutz’ (1925) discussion suggests that an unstressed 1pl pronoun ma may not belong to the traditional repertoire of Alemannic varieties in Vorarlberg. While our participants’ responses from Alemannic Raggal (see map in Figure 3) show both mer ‘we’ and mir ‘we’, in our analysis, we interpret instances of mer as instances of non-CA, but merely as unstressed or clitic pronouns. This is because to our knowledge, unlike with the Bavarian morphemes ma and mia, no documentation exists of a co-occurrence of both mer and mir from Alemannic varieties that would provide a grounds to interpret mer as a CA-morpheme. The two ambiguous instances of 1pl CA in Alemannic Raggal therefore only represent instances where speakers used the form ma. By labeling these as ambiguous CA, to some extent we apply a Bavarian classification to an Alemannic variety, which may not be appropriate from within the logic of Alemannic. As it is, the Alemannic variety again stands out against the Bavarian varieties.

Returning to CA, the mentioned single non-ambiguous instance with 1pl occurs in Oberwölz in the South Bavarian area, but this one (3.8%) instance occurs among several instances of ambiguous (53.8%) and non-CA (42.3%). While the general observed ambiguity in the 1pl makes it hard to draw conclusions, it is nonetheless worth noting its occurrence in comparison to the pattern found with the 2sg. Among the locations in the Bavarian dialect area, the South Bavarian varieties exhibited the lowest degree of CA or even a total absence. While among these South Bavarian varieties Oberwölz showed the most instances, it nonetheless ranked behind the locations in the Central Bavarian or the transitional area. This suggest that the occurrence of 2sg and 1pl CA are not necessarily connected to each in the sense that ubiquitous occurrence of 2sg is a precondition for 1pl CA, but instead grammaticalize independently of each other. The following data from the 2pl and from the open conversations will further substantiate this observation.

The data from the 2pl again show a different pattern (see Figure 11 and Table 8). Connecting to the discussion of Alemannic, Raggal stands out from the other locations in that there are no responses that include CA or ambiguous forms. The other 12 locations show different patterns. In the South Bavarian area, both Tux and Oberwölz show ambiguous (Tux 7.1%, Oberwölz 18.2%) and unambiguous (Tux 28.6%, Oberwölz 22.7%) instances of CA in about a third of the responses. South Bavarian Weißbriach, in contrast, only shows an isolated instance (5.6%) of ambiguous CA next to a large majority of non-CA (94.4%), suggesting that 2pl CA is not a part of the Weißbriach dialect. The South Bavarian
area thus is again not homogeneous. Unlike in the 2sg, where Tux was the only South Bavarian location that did not show any clear instance of CA, with the 2pl it is Weißbriach. The occurrence of ambiguous instances in most locations calls for caution in the interpretation of the results. Yet focusing only on the unambiguous instances, the South Bavarian locations appear to show CA to a lesser extent than those in the Central Bavarian and in the South-Central Bavarian transition area. In Allentsteig, where we find lowest frequency of CA in the Central Bavarian area, nearly half (46.4%) of the responses show unambiguous CA. The South Bavarian locations Tux and Oberwölz do not reach this level even if we did count ambiguous instances as evidence of CA (Tux 35.7%, Oberwölz 40.9%).

Readings of the variation within the Central Bavarian area and the transition area, on the other hand, to some extent depends on the interpretation of ambiguous instances. While ambiguous CA makes up a third (33.3%) of the responses in Taufkirchen (Central Bavarian), they are absent in Passail (South-Central Bavarian transition area) and Tarrenz (Alemannic-Bavarian transition area). Because of the uncertainty that comes from the interpretation of the ambiguous instances, at this point, we forgo a more detailed analysis of the differences in these areas but rather focus on the two locations where there are no ambiguous instances. Both in Passail and Tarrenz, a majority of responses show CA (Passail 60%, Tarrenz 75%). These two locations stand in contrast to the 10 locations that do show ambiguous agreement. We already discussed above how ambiguity arises through the identity...
Table 8. Relative and absolute frequencies of 2pl CA in the LPE-D with C-elements *ob ‘if’, *wann ‘when’, and *wie viele + NP ‘how many + NP’. Data from 13 locations, 144 speakers, \( n = 228 \)

<table>
<thead>
<tr>
<th>Dialect area</th>
<th>Location</th>
<th>CA</th>
<th>ambiguous</th>
<th>no CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Bavarian</td>
<td>Allentsteig</td>
<td>13 (46.6%)</td>
<td>7 (25%)</td>
<td>8 (28.6%)</td>
</tr>
<tr>
<td></td>
<td>Gaweinstal</td>
<td>9 (56.3%)</td>
<td>2 (12.5%)</td>
<td>5 (31.3%)</td>
</tr>
<tr>
<td></td>
<td>Taufkirchen an der Pram</td>
<td>11 (61.1%)</td>
<td>6 (33.3%)</td>
<td>1 (5.6%)</td>
</tr>
<tr>
<td></td>
<td>Neumarkt an der Ybbs</td>
<td>10 (62.5%)</td>
<td>2 (12.5%)</td>
<td>4 (25%)</td>
</tr>
<tr>
<td>South-Central Bavarian (Transition Area)</td>
<td>Steyrling</td>
<td>8 (53.3%)</td>
<td>4 (26.7%)</td>
<td>3 (20%)</td>
</tr>
<tr>
<td></td>
<td>Neckenmarkt</td>
<td>10 (62.5%)</td>
<td>2 (12.5%)</td>
<td>4 (25%)</td>
</tr>
<tr>
<td></td>
<td>Passail</td>
<td>9 (60%)</td>
<td>0 (0%)</td>
<td>6 (40%)</td>
</tr>
<tr>
<td></td>
<td>Hüttenschlag</td>
<td>7 (58.3%)</td>
<td>3 (25%)</td>
<td>2 (16.7%)</td>
</tr>
<tr>
<td>South Bavarian</td>
<td>Oberwölz</td>
<td>5 (22.7%)</td>
<td>4 (18.7%)</td>
<td>13 (59.1%)</td>
</tr>
<tr>
<td></td>
<td>Weißbrach</td>
<td>0 (0%)</td>
<td>1 (5.6%)</td>
<td>17 (94.1%)</td>
</tr>
<tr>
<td></td>
<td>Tux</td>
<td>4 (28.6%)</td>
<td>1 (7.1%)</td>
<td>9 (64.3%)</td>
</tr>
<tr>
<td>Alemannic-Bavarian Transition Area</td>
<td>Tarrenz</td>
<td>18 (75%)</td>
<td>0 (%)</td>
<td>6 (25%)</td>
</tr>
<tr>
<td>Alemannic</td>
<td>Raggal</td>
<td>0 (%)</td>
<td>0 (%)</td>
<td>14 (100%)</td>
</tr>
</tbody>
</table>

of the 2pl clitic pronoun -s (< Bavarian es ‘you’) and the CA-morpheme -s. In this regard, it is worth pointing out differences between Passail and Tarrenz. Passail generally matches the pattern of 2pl CA discussed so far in that the CA-morpheme is -s. We will return to this shortly but first focus on Tarrenz. As already pointed out earlier, the CA morpheme in the dialect of Tarrenz is generally -ets, as can be seen in (16).\(^{15}\)

(16) CA with 2pl (dialect example from Tarrenz, Tyrol, Austria, in the Alemannic-Bavarian transition area, DiÖ Corpus)

[...], *ob-ets es morgen arbeit-ets
[...], if-2pl 2pl tomorrow work-2pl
‘if you work tomorrow’

\(^{15}\) The parallel appearance of the pronoun es ‘you’ backs the interpretation of -ets as a CA-morpheme. A further possibility, the existence of a pronoun etses, is not support by further data. In a different experiment task, the same participant e.g. used es without an -ets morpheme: ‘wie long es morge schlafets’ (‘how long you sleep tomorrow’).
To our knowledge, this form of 2PL CA has not been previously been documented. We want to briefly discuss different paths that may explain its emergence. A first path to explaining the -ets-morpheme may be pursued when looking at the parallel morphemes of CA and verbal inflection in (16). As both morphemes are identical, it is conceivable that the ending is a product of copying the verb inflectional ending (as it appears in inversion contexts) to the C-element (as described by Hoekstra & Smits 1998). While this accounts for the identity in form of both morphemes, it is not an entirely satisfactory explanation, as it raises the question of how the verb inflectional morpheme -ets itself emerged.

We argue that plausible explanations rely on clitization of pronouns and reanalysis, following the path proposed by Weiß (2018). Two related paths may explain the emergence of the verb inflectional ending: The first explanation assumes the emergence of a new verb inflectional morpheme. The inflectional morpheme -ets may itself be the product of reanalysis of a verb ending -et, which corresponds to endings for the 2PL in earlier stages of German (Braune & Heidermanns 2018: 360). Contexts where this -et ending is followed by a clitic pronoun -s could have provided a foundation for reanalysis, leading to an inflectional ending -ets. This clitization and reanalysis would have provided the entry for the reanalysis process proposed by Weiß (2018) described above, which first led to a stage of DA that provided the foundation for the association of the inflectional ending with the C-position, i.e. CA. As the verb ending also appears in verb final clauses, a subsequent generalization must have taken place to reach the current state. This explanation accounts for both the verb inflectional ending and CA at the same time.

The second explanation assumes an immediate emergence from a pronoun ets, analogous to the other instances of CA discussed here. However, there is no direct evidence of this, as our data does not indicate a pronoun ets in the contemporary dialect of Tarrenz. However, 2PL pronouns appear in similar forms in other Bavarian varieties, e.g. äds (Renn & König 2006: 86). While this similar form points to a general credibility to the pronominal origin of the morpheme, it is to some degree problematic as evidence, because this pronoun is restricted to a rather small area North of Regensburg (in Southern Germany), and thus more than two hundred kilometers from Tarrenz. This evidence can however be substantiated by further

---

16. This in itself is not problematic but rather supported by a theory of progressing CA grammaticalization. With an increasing grammaticalization of a pronoun ets both as a verbal agreement- and CA-morpheme, its pronominal character may have become increasingly less prominent to speakers and thus paved the way for the use of es as a personal pronoun, which could have been ben adopted e.g. in contact with other varieties, leading to the present state where ets does no longer appear as a pronoun but only as agreement morphology.
data from our project’s corpus from Tux, which is located markedly closer in about 100 kilometers distance East of Tarrenz. (17) gives an example of a 2PL pronoun its from Tux. Notwithstanding the difference in the initial vowel, the form is very similar and can support the proposed pronominal origin of the -ets CA-morpheme in Tarrenz.17 These pronouns ets or its may themselves have been the product of reanalysis processes of verb inflection and pronouns, as documented in other instances (Weiß 2018: 149). This new pronoun may then have entered the path to CA via clitization and DA as outlined above.

(17) CA with 2PL (dialect example from Tux, Tyrol, Austria, South Bavarian, DiÖ Corpus)
Wisst its denn scho, wenn-/=s mer wieder kämm-ts?
know 2PL MODAL-PARTICLE already when-2PL/=you me back come-2PL
‘Do you already know when you’ll be back?’

From this discussion of potential sources of the 2PL CA-morpheme -ets, it appears that an account that assumes a straightforward adoption of the verb inflectional ending in the C-position falls short of explaining its ultimate emergence, as it does not capture where the verb ending itself originated. Explanations that assume a reanalysis of verb inflection and pronouns, on the other hand, explain both the emergence of CA and of new verb inflection, which in turn provide them more credibility until further evidence is presented.

As a further point of discussion regarding Tarrenz, it is worth noting an exception to the general pattern of CA-morphology. (18) shows the use of an -s-morpheme for 2PL CA by a young speaker from Tarrenz, who stands against a majority of 9 speakers showing 2PL CA with the -ets variant. This, as with the occurrence of 2SG CA in Weißbriach (see Figure 9 above), may be an indication of change in the dialect of Tarrenz towards a form that is common in Eastern Austria.

17. It is, of course, worth noting that the CA-morpheme in this example is not -ets (or -its). Rather, this example shows an instance of ambiguous agreement with the -s-morpheme. This does not undermine the hypothesis that a pronoun ets (or a related form like its) may have been the origin for the CA-morphology in Tarrenz. Rather, we think it can illustrate that the grammaticalization of agreement morphology can take different paths. Different competing forms of personal pronouns may take different functions in different syntactic contexts and subsequently take different paths of grammaticalization. Although we propose a shared origin for the different morphemes discussed here (pronouns ets and its and äds as well as CA- and verbal inflectional morphem), we opt for different spellings here, as the phonological differences between the morphemes of Tarrenz and Tux in our corpus appear markedly different in regards to the initial vowel and are not voiceless as suggested by the äds-spelling found in the literature. Without further evidence, we refrain from conflating these morphemes into one.
We now turn to the absence of ambiguous instances of CA in Passail. As noted, Passail matches the wider Bavarian pattern of a 2PL CA morpheme -s. Unlike the other Bavarian locations investigated, in Passail there are only unambiguous (60%) CA and instances of non-CA (40%). The lack of ambiguity is the result of a response behavior of the participants from Passail. Unlike in other investigated locations, all responses that show a -s-morpheme also show a full personal pronoun. In all instances, this is the pronoun ihr ‘you’ (that is considered the 2PL pronoun in standard German), but never es, which immediately connects to the cliticized -s form, as shown in (19). This is worth remarking upon because it sets apart Passail from the other Bavarian locations, where the pronoun es (or the related dialect form des ‘you’) occur to different extent. This raises the question of whether this absence of a pronoun es and the seemingly obligatory use of the full pronoun ihr ‘you.2PL’ in the context of CA are interconnected. At this point, we can only speculate on a potential causality or directionality in such a grammaticalization process, and the details of the interaction of pronominal system and CA require further scrutiny. Yet it appears plausible that the appearance of the standard pronoun ihr ‘you’ itself, as well as what appears to be its obligatory use in combination with CA, indicates a bleaching of the pronominal nature of the -s-morpheme in the dialects. 2PL CA in Passail may have reached a higher level of grammaticalization compared to other locations where this morpheme appears to require the use of a pronoun to a lesser extent.

A resource that may help in assessing the validity of such a theory is data from the so-called Wenker sentences, a survey of dialects conducted in the late 19th and early 20th century. Despite methodological issues of the survey, they provide data for much of the (former) German-speaking area. Within the immediate vicinity of Passail, three locations, Arzberg, Fladnitz bei Passail, and St. Kathrein am Offenseck, were surveyed in the first half of the 20th century and thus allow for insights into earlier stages of the pronominal system in the region that may inform our view of the present state. At the heart of the historic survey was the translation of
from Standard German into the local dialect. Of the 40 sentences used in the survey in Austria, four (sentences 28, 30, 31, and 32) include a 2PL subject pronoun (none of which appear in subordinate clauses, unfortunately).

The three locations indicate the following use of 2PL subject pronouns. In Fladnitz, none of the sentences show any use of a 2PL pronoun, but rather indicate pro drop and, in one sentence, the use of an imperative form. This is remarkable, as it suggests that pro-drop historically may have been mandatory. Data from the other locations, however, does not substantiate this suggestion. In Arzberg, two sentences use dös as pronoun, while one sentences uses ös and the final sentence appears to show a 2SG subject with pro-drop. In the last location, St. Kathrein, three sentences show the pronoun dös, the fourth sentence 2PL pro-drop. Both dös and ös can be considered alternative forms of es and are compatible with the theory of clitization. The historical data therefore suggest that ihr replaced ös or des in the last century, but CA remained part of the variety despite no longer being backed by the morphological similarity of pronoun and CA-morpheme.

In sum, the data shows that there are markedly different patterns for the three different inflection forms of CA. The (almost) complete absence of unambiguous evidence for 1PL CA suggests that it is the least widely spread of the three forms. 2SG CA on the other hand is widely spread but appeared limited to Eastern and Central Austria. Areal distribution is widest for 2PL CA. These findings fit the borders of the dialect areas to different degrees. Occurrence of the 2SG forms very much aligns with the dialect boundaries and separates the South Bavarian varieties from its northern neighbors in the Central Bavarian area. In contrast to this, for the 2PL, the three locations in the South Bavarian area showed variation in that only two of the locations evidenced CA. The transition area between South Bavarian and Alemannic stood out because of its previously undocumented form of 2PL CA. Overall, the lack of clear evidence of CA in the Alemannic location stood out against the general Bavarian area, suggesting that the grammatical difference between these regions is indeed substantial regarding this grammatical aspect. This speaks to a typological difference between Alemannic and Bavarian concerning CA that we here only want to note, but will not try to account for.

The different patterns of CA let us draw different conclusions on the areality of CA. On the one hand, the contrast between Alemannic and Bavarian suggests that Bavarian is “better suited” to show CA than Alemannic. Within the Bavarian dialect area, there are further differences. Previous research on 2PL CA by Lenz et al. (2014) had indicated that it is essentially limited to Central and North-Bavarian varieties and occurred only marginally in the South Bavarian area. Although our findings show evidence of 2SG (Oberwölz) and 2PL (Oberwölz, Tux) CA in South Bavarian varieties, it remains less frequent there compared to the Central Bavarian varieties. Therefore, the Central Bavarian area appears as the
center of CA in Austria when considering the 2sg and 2pl forms. Our data thus both confirms and refines previous findings. Unambiguous evidence of 1pl, on the other hand, while very scant even there, only comes from South Bavarian. This in turn means that in those areas, where 2sg and 2pl forms are most frequent, there is no clear evidence of 1pl CA. The only location where all three forms of CA appeared in the experiment – data from conservations will be considered shortly – is Oberwölz, which is located in the South Bavarian dialect area. However, Oberwölz shows 2sg and 2pl forms less frequently than the Central Bavarian locations. The location where the paradigm is most complete thus appears not to be the location where CA is most obligatory. We may thus consider the different subject forms of CA to as to some extent independent, in that abundance of 2sg and 2pl CA does not provide the foundation for 1pl CA. The absence of 2sg CA from Western Austria substantiates this point and, at least for Austria, calls into question the hierarchy that assumes the 2sg as the minimal form.

We will revisit areality to discuss the historical emergence of CA shortly. Before doing so, we briefly consider the data collected through the LPEs in light of more “natural” data from recordings of conversations.

### 3.4 Analyses on ‘natural data’: CA variation in “conversations among friends”

The discussion has so far focused on data elicited through LPEs. As previously stated, this experimental approach aims at the elicitation of the participants’ dialect or Standard German register. The language speakers use in the experiment may to a certain (or considerable) extent be idealized. One of the central questions of the SFB concerns how this language use “in vitro” relates to the speakers’ language “in vivo”, i.e. in what represents more natural communicative situations. To illustrate this, we briefly discuss first findings from our conversations between friends from selected locations. As already mentioned, these conversations between two speakers are not monitored by an interviewer. The field worker only sets up the microphones and recording equipment and instructs the participants on the use of cue cards. These cards are intended to guide the conversations and touch a number of topics, e.g. other varieties of German or language change, that are relevant to language attitudinal research performed in other project parts. Thus, this setting not only provides insight into what probably represents a more informal linguistic register compared to that used in the interview, but also allows the investigation of language attitudes in a different context. As the linguistic register used, these attitudes may be different from those voiced in the interview, as speakers may e.g. be reluctant to voice certain opinions with an unfamiliar person.
Table 9. Absolute frequencies of C-elements in open conversations for 16 speakers from 3 locations

<table>
<thead>
<tr>
<th>C-element</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2SG</td>
</tr>
<tr>
<td>ob ‘if’</td>
<td>7</td>
</tr>
<tr>
<td>wann ‘when’</td>
<td>1</td>
</tr>
<tr>
<td>wenn ‘if, when’</td>
<td>80</td>
</tr>
<tr>
<td>wie viele + NP ‘how many + NP’</td>
<td>0</td>
</tr>
</tbody>
</table>

In the following section, we discuss data taken from conversations recorded with 16 speakers in four different locations. Each location is represented by two older and two younger speakers. All speakers analyzed here also participated in the LPEs. Neckenmarkt and Neumarkt represent the area we identified as the area that shows 2sg and 2pl CA the most. Raggal, on the other hand, did not show any indication of CA in 2sg and 2pl. Weißenbruch, finally, showed only limited evidence of CA in the experiment, but is located in the South Bavarian dialect area, where evidence of 1pl CA was found. Each of the conversations is about one hour in length. As can be seen in Table 9, the C-elements targeted in the experiments occur only rarely in the conversations, whereas wenn ‘if, when’ occurs more frequent. The dominance of 2sg subjects and the almost complete absence of 2pl equivalents is rooted in the communicative context where two speakers interact with each other. In the following discussion, we assume that wenn behaves like ob and wann and discuss the occurrence of CA without further distinction. Because of the scarce evidence, we omit discussion of the 2pl.

Figure 12 shows the regional occurrence of 2sg CA and indicates that the data collected by the experiments very much reflects the appearance in unmonitored conversations. While the absolute numbers of sentences that might exhibit CA varies strongly between speakers, there is remarkable consistency in the data in that speakers in Neumarkt and Neckenmarkt exhibit CA without exception, whereas it is consistently absent with speakers in Raggal and Weißenbruch. Based on this data, 2sg CA appears obligatory.

We can contrast this with the speakers’ responses in the dialect LPE targeting 2sg represented in Figure 13. The comparison of speakers’ responses shows that two speakers (213 and 263) did not (consistently) use CA in the experiment. Two speakers from Raggal elude comparison as they did not provide the targeted responses in the LPE. This lack of CA with the Eastern speakers could indicate that CA ultimately is not as obligatory as the data from the conversations suggested. However, it could also indicate that there is some aspect of the LPE’s setup...
that prevents speakers from employing CA. To better understand this pattern, further methodological inquiry is required. We will not pursue this question any further but hope to return to it in future work.

Data for the 1PL, as shown in Figure 14, is much scarcer in the conversations. For five speakers, there are no contexts that could show CA in the recording. Nonetheless, it allows some relevant observations. The data for Neckenmarkt and Neumarkt mirror the observations of ambiguous 1PL CA in these locations.
The relatively frequent occurrence of ambiguous instances with one speaker in Raggal is somewhat noteworthy considering the rarity of such occurrences in the experiment. Nonetheless, it does not contradict the general observation of the experiment that CA is absent in Alemannic speaking Austria. In Weißbriach, however, there is a combined occurrence of the -ma-morpheme and a personal pronoun, indicating 1PL CA. While this evidence is scarce and relies on a single instance (as was the case in the experiment for Oberwölz), the relative proximity to Oberwölz in the South Bavarian area and the occurrence with an older speaker give some plausibility to the observation. This occurrence in Weißbriach casts the South Bavarian area as the general area for 1PL CA. Considering the limited presence of other forms of CA in this area, again supports the assumption that different forms of CA may have emerged independently. The overall scarcity of 1PL CA and the fact that it occurs with an older speaker could indicate that 1PL in general is a feature that is sensitive to decay and only used by a minority of speakers and/or not necessarily consistently so.

3.5 The emergence of CA paradigms in Austrian varieties of German

The data presented so far gives an empirical foundation for a discussion of the theory of the development of CA, particularly as proposed by Weiß (2018). As discussed above, the theory assumes that CA is dependent on clitization of pronominal elements, and that non-pronominal forms of CA can only emerge once such a pronominal “blueprint” for CA has been established. The data here are limited to three forms of a paradigm that can contain up to six forms, and thus cannot be considered complete. Nonetheless, they allow for some observations.
First, the documented forms of 2SG, 1PL, and 2PL CA all appeared compatible with a theory that assumes a pronominal origin of CA. We did not find any convincing evidence for forms of verbal CA that emerged independent of pronominal CA. This speaks to the general soundness of a theory that assumes pronominal CA is the foundation for all further CA, at least for the Bavarian varieties investigated here. However, this does not mean that the observed instances of CA are homogeneous in their emergence. In the cases of 2SG -st and 2PL -ets, we can assume reanalysis of both verb inflection and pronouns resulting in CA and new verb morphology. For 1PL -ma, the clitic pronoun as such appears as the CA morpheme. In the instance of 2PL -s, the CA-morpheme also is the clitic pronoun yet differs from the verb inflectional morpheme -ts, as seen in (17) above. Particularly this last instance may be worth future scrutiny. The morphological difference does not constitute DA, as the verb inflection is not dependent on verb placement, but rather appears as -ts in all positions. This is in sharp contrast to the cases of 2SG -st and 2PL -ets, where the reanalyzed forms appear uniformly as verb inflection and CA-morphemes. In its present state, the theory does not account for these differences. Future work should also address the question of whether this grammaticalization path can be applied equally well to other Germanic varieties.

The regional patterns of the CA paradigms uncovered through the survey further enrich our knowledge of the emergence of CA in one further aspect. We found only one location (Oberwölz) where the three targeted forms of CA occur parallel, whereas two locations (Tarrenz, Tux) showed CA only in the 2PL. One further location (Weißbriach) data suggests that 1PL CA may have traditionally existed independently of 2SG and 2PL, although the data is sparse. What is missing, in fact, is an occurrence of only 2SG CA. The 2SG therefore does not appear as the fundamental form of CA that is necessary for the development of other forms. Assuming that no forms of CA disappeared after their emergence, this suggests that CA may have emerged independently multiple times following a pattern of clitic reanalysis.

4. Summary and conclusion

This contribution has provided a multidimensional study of CA in varieties of German in Austria. Using different elicitation methods, we set out to investigate the interplay of linguistic and situational factors and the geographical distribution of CA. Through our experimental approach (LPE), we provide the first study of CA to focus on a corpus of spoken language, as opposed to questionnaires or user judgements. It is also the first study to examine the place of CA in speakers’ registers (dialect versus standard). The experimental methodology,
conducted using speakers from 13 locations in Austria, proved successful and supported multiple new insights into the phenomenon. The experiments elicit CA with C-elements that are rare in open conversations and thus allow us the targeted investigation of linguistic factors that control its occurrence and variation. This highlights the methodological value of LPEs. Our examination of speakers’ registers could show that speakers do not employ CA in spoken Standard German, marking it as a feature of non-standard varieties of German in Austria. While we only discussed standard language data from a subsample of our research locations, the complete absence of CA in this data, even in locations where it appears frequently in the dialect, suggests a strong association of CA with non-standard varieties of German.

Our analysis showed numerous linguistic patterns that influence the occurrence of CA. We were able to provide strong evidence that CA is dependent on the complexity of the elements in C-position: Less complex elements, such as subordinating conjunctions (like ob ‘if’) and (morphologically) simple wh-words (like wann ‘when’), show CA much more frequently than wh-phrases (like wie viele + NP ‘how many + NP’). Compared to the related phenomenon DFC, this provides an essentially inversed hierarchy for the (Bavarian) varieties that were investigated. The elements least likely to show CA are the most likely to show DFC. The wh-element wann ‘when’ takes a middle ground in that both DFC and CA appear, although DFC remains rare.

The linguistic factors, in part, connected closely to areality. Across all locations, CA showed the greatest areal distribution with 2pl subjects, followed by 2sg forms. While ambiguous instances of 1pl CA appeared in all locations, and frequently in most cases, clear evidence only occurred with individual speakers. This suggested that Austria, when considered as a linguistic space, much like Hesse, does not follow the generalization encountered in the literature that the 2sg is the most frequent form of CA.

We further used the data to evaluate a theory of the grammaticalization of CA from clitic pronouns as proposed by Weiß (2018). Our data proved compatible with such a theory yet also raised the question of how the differences between CA and verb inflectional morphology can be explained. Through our survey, we uncovered a previously undescribed form of 2pl CA in Tarrenz but also an instance where the 2pl pronoun es that appears as the historic source of CA appears to have been replaced by ihr, while a 2pl CA morpheme -s has remained part of the variety, but only occurs in combination with the full pronoun. These instances suggest that CA can emerge independently and may be at different stages of grammaticalization. Our synchronic variationist data thus also allows for new insights into the diachrony of CA.
The investigation of areality went beyond questions of general diffusion and frequency of CA with different sentence subjects, as we considered our findings exploring dialect geographical work. While in some instances, this variation follows the borders drawn by earlier dialectology (guided by phonology), in other instances these borders did not capture the encountered patterns. The division between Alemannic and Bavarian varieties also appeared to mark a border for the occurrence of CA. While there was evidence of CA in all Bavarian locations investigated, the Alemannic location stood out as showing no evidence of CA. CA thus mirrors the existing distinction between Bavarian and Alemannic varieties. The pattern drawn in the different Bavarian subregions proved more intricate. Generally, the Central Bavarian dialect area and the Central-South Bavarian transition appeared as the center of CA in Austria, showing both 2SG and 2PL forms. The South Bavarian locations we investigated appeared more heterogeneous. While 2PL CA appeared in both the westernmost (Tux) and the easternmost (Oberwölz) South Bavarian location, there was no clear evidence in the central location (Weißbrach). However, while 2SG CA appeared in eastern and central (albeit potentially as a recent innovation due to speaker mobility) South Bavarian, there was no evidence in the western South Bavarian area. The appearance of isolated evidence of 1PL CA in two South Bavarian locations further adds to the heterogeneity of this region regarding CA. The westernmost South Bavarian location (Tux), with its absence of 2SG CA and presence of 2PL CA, mirrored the location in the Alemannic-Bavarian transition area (Tarrenz). However, both locations continued to demonstrate a marked difference, since, as already mentioned above, Tarrenz showed the previously undescribed agreement morpheme -ets. These data go against descriptions of the 2SG as a minimal form but suggest that the 2PL may be just as suitable a starting point for CA, as might the 1PL, although the areal distribution appears much more limited.

The investigation of speaker registers through the LPEs could identify CA as a feature of non-standard varieties. While the comparison of the experiment data to that from the open conversations generally confirmed this, it also raised methodological questions. The most prominent methodological question is: to what extent do the LPE-D elicit speakers’ competence – a question that our new elicitation method has to reflect on, just like any other established method. Our examination of selected speakers’ linguistic behavior in conversations with other locals could provide first insights into this question. The discussion was limited to a small number of speakers, yet nonetheless, this data provided further important insights. As the conversational data was, at best, patchy for 1PL and 2PL subjects, it illustrated the relevance of the experimental approach for systematic elicitation of data on specific features. The collected conversational data proved largely compatible with the data from the experiments. However, in the case of 1PL CA, they
also indicated the value of a combination of methods to capture such apparently rare phenomena.

While this combined methodology opens our data to questions, we maintain that the more general findings on linguistic factors and areal patterns are robust and provide an unprecedented account of CA in varieties of German in Austria. However, it also underscores the importance of further work into how different methodologies for the investigation of particularly syntactic variation influence research outcomes. We hope to build on the findings presented here and to pursue the open questions in further work.

**Funding**

Research for this contribution was conducted in the SFB “German in Austria. Variation – Contact – Perception” (FWF 060, principal investigator: Alexandra N. Lenz), funded by the Austrian Science Fund.

**Acknowledgements**

The research for this contribution benefited from the help of several colleagues. Special mention deserve Ludwig Maximilian Breuer, Pamela Goryczka, Melanie Seltmann, and Anja Wittibschlager for their support in data processing and visualization. We further thank Ellen Brandner, Oliver Schallert, Helmut Weiß, and an anonymous reviewer for their comments on earlier drafts of this paper.

**References**


**Address for correspondence**

Matthias Fingerhuth
University of Duisburg-Essen
University Library
Universitätsstraße 9
45141 Essen
Germany
matthias.fingerhuth@uni-due.de
Co-author information

Alexandra N. Lenz
University of Vienna
Department of German Studies
alexandra.lenz@univie.ac.at