Unifying species of C-agreement*

Thomas McFadden & Sandhya Sundaresan

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), University of Göttingen

1. Introduction

Work within the so-called cartographic enterprise (see e.g. Rizzi 1997, Cinque 1999) has presented a range of crosslinguistically robust evidence to argue that the clausal periphery is more functionally articulated than was previously assumed, with a sequence of functional C heads in a more or less fixed order. The central goal of this paper is to provide further evidence for this from a relatively new empirical landscape, namely the realm of complementizer- or C-agreement across languages and structures. The templatic proposal that we develop here represents one of the first attempts (but see also Baker 2018) to unify the full range of C-agreement patterns across languages and clause-types. Such evidence shows that the heterogeneity of C-agreement can be naturally derived by positing that each such type of agreement involves a dedicated, distinctly ordered C head, paralleling the idea (going back to the "split Infl" proposal in Pollock 1989) that differences in clause-internal agreement are indicative of the involvement of multiple functional heads (e.g. T, v) within the clause.

2. Varieties of C-agreement

We can distinguish at least three types of C-agreement: downward complementizer agreement (DCA), upward complementizer agreement (UCA), and allocutive agreement (AA).

DCA is characterized by the following empirical signature: the embedded C head shows agreement with its clausemate embedded subject, as in (1):

(1)
$$\operatorname{Subj}_{Matrix} \ldots \operatorname{C} \ldots \operatorname{Subj}_{Embedded} \ldots$$

DCA is shown for Flemish (2) where the embedded complementizer *dan* ('that') reflects 3PL agreement with the embedded subject *ze* ('they'):

^{*}We gratefully acknowledge the audiences at NELS 51 (UQAM, Montreal) and, more recently, at the ComSyn colloquium series (Leiden) for useful feedback on versions of this idea. All errors remain our own.

(2) K peinzen da-n ze morgen goan.

I think that-3PL they tomorrow go-3PL
'I think that they will go tomorrow.' (West Flemish, Haegeman 1992)

In cases of UCA, an embedded C head shows agreement with the subject of a higher clause, as templatically shown in (3):

(3)
$$\operatorname{Subj}_{Matrix} \ldots \operatorname{C} \ldots \operatorname{Subj}_{Embedded} \ldots$$

UCA has been reported in the literature for Bantu languages like Limbum (Nformi 2018) and Lubukusu (see e.g. Diercks 2013) below. In Lubukusu (4), the embedded complementizer *bali* ('that') reflects Class 2 agreement with the matrix subject *babandu* ('people'):

(4) Ba-ba-ndu ba-bol-el-a Alfredi ba-li a-kha-khil-e.
2-2-people 2-said-AP-FV 1Alfred 2-that 1-FUT-conquer
"The people told Alfred that he will win." (Lubukusu, Diercks 2013)

The characteristic signature of AA is that it involves an embedded or matrix C head showing agreement with the Addressee. AA thus has the templatic structure in (5):

(5)
$$(Subj_{Matrix}) \dots C \dots Subj_{Embedded} \dots$$
Addressee

In Basque (6), the suffix -*n* on the complementizer *din* reflects the fact that the utterance is addressed to a close female friend of the speaker:

(6) Pettek lan egin di-n.
Peter.E work.A do.PRF 3A.3E-ALLF
'Peter worked.' *Uttered to a close female friend* (Basque, Oyharçabal 1993)

Prior analyses of these phenomena vary, with disagreement centering largely on whether they instantiate agreement in the first place, and, assuming they do, on how such agreement should be formally derived. For instance, van Alem (2020) argues that DCA involves clitic-doubling rather than agreement, Weisser (2019) that it reflects allomorphy and Fuß (2008) and Ackema and Neeleman (2004) that it results from operations at the syntax-PF interface. Within the agreement approach, analyses differ with respect to whether DCA is due to T-to-C movement (Zwart 1997), downward probing from C (Carstens 2003, van Koppen 2005) or upward probing following subject-movement to a position above the C probe (Zeijlstra 2012). Similarly, there is disagreement in whether UCA is a form of agreement or reflects anaphora due to Agree between a C head (Diercks 2013) or a speech-verb (Driemel and Kouneli 2020) and a silent clause-peripheral anaphor. Agreement approaches to UCA also vary in whether they treat it as a form of upward (Baker 2008, Carstens 2016) or downward (following covert C-movement to a position higher than the goal) C-probing (Diercks et al. 2020). Turning to AA, Oyharçabal (1993) analyzes this as C-agreement with

an adjunct. Alternatively, such agreement is argued to be with the syntactic representation of the addressee in the Speech Act Phrase (SAP) high in the clausal periphery, either via downward C-probing following C-movement (Miyagawa 2017) or upward C-probing (McFadden 2020).

The goal of this paper is not to adjudicate the choice between the individual analyses above, but to develop a unified analysis that can accommodate the patterns of different types of C-agreement in comparison. While DCA and UCA have been comparatively assessed (see the overview in van Koppen 2017), AA is typically not included in the mix. This paper seeks to redress this gap. Going forward, we will leave open the possibility that what looks like C-agreement in some languages may actually require a different type of analysis (e.g. in terms of interactions with anaphora, clitic-doubling or PF operations). At the same time, we will adopt the working assumption that an Agree-based account involving a C probe is correct for at least *some* instances of each of the three types and restrict ourselves on deriving these in comparison.

3. Articulating the puzzle

The fundamental challenge of C-agreement is its heterogeneity which resists a straightforward analytic unification. DCA, UCA and AA arguably differ along two independent parameters: whether the goal is argumental, and whether C-probing is upward or downward:

		Goal	
(7)	AGREE DIRECTION	ARGUMENT	Non-Argument
	Downward	DCA	_
	Upward	UCA	AA

The simplest imaginable unified analysis, involving a ϕ -probe on a single C head, is not nuanced enough to capture the diversity of patterns schematized above. Even if we had a mechanism to derive the directional difference between UCA and DCA, it is difficult to see how we might also derive AA without relativizing the C-probe, essentially by brute force, to target the addresseee (rather than a clausal argument).

4. Toward a solution: two observations

The solution we develop to this intriguing challenge is informed by two independent observations concerning the different types of C-agreement.

4.1 Observation 1: {D/U}CA vs. AA

Observation 1: DCA and UCA are restricted to structures that exclusively instantiate the properties of embedded clauses while AA is, in contrast, a root phenomenon. For instance,

C-agreement in Frisian is disallowed in embedded clauses just in case these are verb-second, which is a property of root clauses (de Haan 2001), cf. (8) vs. (9): ¹

- (8) Heit sei dat-st do soks net leauwe moa-st. dad said that-2sg you such not believe must-2sg
- (9) Heit sei dat *-st do moa-st soks net leauwe. dad said that-2sg you must-2sg such not believe 'Dad said that you should not believe such things.'

The behavior of AA contrasts strikingly with this. AA is, first of all, famously impossible in embedded clauses in Basque (and several other languages Oyharçabal 1993, Antonov 2015). Thus, root AA in (10) is grammatical while embedded AA in (11) is not:

- (10) Ez din at nahi [gerta dakion]. not 1E.ALLF want happen 3A.3D.C 'I don't want it to happen to him.'
- (11) *Ez dinat nahi [gerta dakion an]. not 1E.ALLF want happen 3A.3D.ALLF.C 'I don't want it to happen to him.'

Second, even in languages like Japanese & Tamil where embedded AA *is* attested, it is restricted to complements of typical bridge verbs (as in Tamil (12)) and other types of embedded clauses which independently display the syntactic hallmarks of root CPs (Miyagawa 2012, McFadden 2020):

Maya [avæ pootti-læ ˈdejkkæ-poo-r-aal-ngæ-nnŭ] so-nn-aa.

Maya [she contest-LOC win-go-PRES-3SF-ALLOC-COMP] say-PST-3SF

'Maya said that she would win the contest.' (to honorific Addressee)

4.2 Observation 2: DCA vs. UCA

Observation 2: UCA must involve a higher C head than DCA. First, UCA is commonly associated with interpretive effects pertaining to the speaker (Diercks 2013, Diercks et al. 2020). For instance, Diercks (2013) notes that *a-li* in Lubukusu (13) is only possible if the speaker considers the reported information reliable; otherwise non-agreeing *bali* appears.

(13) Mosesi a-lom-ile a-li Sammy k-eb-ile chi-rupia.

1Moses 1-say-PRF 1-C 1Sammy 1-steal-PST 10-money
'Moses has said that Sammy stole the money.'

¹This is not (just) about the presence of an overt complementizer. For instance, in Bavarian, C-agreement can attach to a *wh*-phrase in Spec-CP in an embedded clause with no overt C (Bayer 1984).

Such data also show that subject UCA in a language like Kipsigis tracks the source of information (13), indicating the involvement of a structurally higher C head, like a hearsay evidential (Speas 2004).

Second, UCA complementizers often show a closer morphosyntactic connection to the matrix clause, frequently being grammaticalized from 'say' verbs and even being able to replace the matrix verb in its entirety (as shown for Kipsigis, Diercks et al. 2020).

Crucially, such patterns have not been reported for DCA, suggesting that this involves a lower C, more closely associated with the embedded clause.

5. Proposal

The proposal we develop here derives the heterogeneity of C-agreement by tweaking the following parameters across individual languages and structures: (i) the availability of a given C head and its ability to host a ϕ -probe; (ii) the height of such a probe relative to the CP phase; and (iii) the availability of a ϕ -goal within the C domain.

5.1 Deriving DCA vs. UCA

We propose that DCA and UCA involve distinct ϕ -probes on Fin & Force, respectively, with Force being merged hierarchically higher than Fin (Observation 2). Furthermore, we assume that Fin is merged below the embedded CP phase while Force is merged above it (see, for discussion and motivation, Carstens 2016). UCA thus diagnoses a more elaborated C-domain, involving at least Force and Fin, yielding the concomitant interpretive consequences discussed in Section 4.2.

Being below the CP phase, Fin can probe downward into the embedded clause. The ϕ -probe on Fin is valued by the minimally closest nominal in its search-space, namely the embedded subject, yielding DCA (15):

(15) **DCA** :
$$[Subj_{Mat}]$$
 $[phase [Fin_{[\phi:]} Subj_{Emb}]]]$

Being merged above the embedded CP phase, Force cannot probe into the embedded CP, (under the PIC, Chomsky 2001). But the ϕ -Probe on Force is still active triggering, we propose, an independent operation of *domain expansion* (independently proposed in e.g. Béjar and Rezac 2009, Clem 2019) whereby the search-space of the ϕ -probe is extended above the Force head. This allows the probe to search upwards in a second cycle of Agree, yielding UCA as agreement with the closest viable goal.

(16) UCA:
$$[Subj_{Mat} [Force_{[\phi:]}]_{phase} [Fin Subj_{Emb}]]]]$$

Note that we still need additional mechanisms to yield targeted Agree with a subject, rather than an object, particularly in cases of UCA. After all, if the matrix clause is transitive, the matrix object, being minimally closer to the ϕ -probe on Force, should bleed UCA with the matrix subject. One solution would be to have a featurally-relativized C probe that distinguishes subject from object, e.g. in terms of case or topic features (in a language where subject and object are reliably distinguished along these lines). Another option might be that the matrix object is simply not a viable goal for the probing C head, e.g. because it is obliquely case-marked or occupies a position that does not c-command the probe. At the end of the day, we believe that the undeniable fact of variation in clausal syntax, both across languages and sentences, forces us toward a solution whose merits can only be decided on a case-by-case basis. But whatever the explanation, it is worth noting that it parallels the nature of subject vs. object asymmetries for clause-internal agreement. A well-known observation is that clause-internal object agreement is typically only possible in a anguage that also allows subject agreement. Interestingly, this also seems to be true for C-agreement: e.g. Kipsigis displays UCA with objects, but only on complementizers that also show agreement with the matrix subject (with the notable caveat that what we are calling UCA might actually involve, not agreement at all, but clitic doubling, as argued in Diercks et al. 2020). This suggests that an explanation of the one might also serve as an explanation for the other.

5.2 Deriving AA

In accordance with Observation 2, we propose that AA involves embedded root clauses which are characterized by the representation of a SpeechActP (SAP), which hosts silent representations of the author & addressee (as independently proposed in Speas and Tenny 2003, Hill 2007, Sundaresan 2012, Krifka 2017). We are assuming that the functional sequence is more or less monotonic: i.e. the presence of a higher head thus typically entails that of a lower one, but not vice-versa. Since the SAP is typically thought to occupy the highest level of the CP, its presence entails the availability of a fully extended C-domain which is only available in (matrix/embedded) root clauses. This immediately explains the comparatively restricted distribution of AA across languages and clause-types.

We will propose here that the AA ϕ -probe is as high as or higher than the UCA ϕ -probe — for now, we will provisionally label the head 'High-C'. As a result, the AA probe is necessarily also above the embedded CP phase, with the result that it cannot search inside the embedded CP phase (due to the PIC). Failed downward Agree + a still active ϕ -probe triggers domain expansion causes the High-C head to probe upward in this new search space. So far, the derivation proceeds exactly as with UCA. The difference between the two emerges at this stage of Agree. Unlike with UCA which involves a less articulated C layer, the addressee in the SAP, in cases of AA, is available as a minimally closer local goal for the upward-probing High-C head. The second-cycle of Agree thus successfully terminates with the addressee in SAP obviating the need for High-C to probe any higher (17):

(17) **AA:**
$$[(Subj_{Mat})]_{SAP}$$
 Addr $[High-C_{[\phi:]}]_{phase}$ $Subj_{Emb}]]]]$

5.3 Deriving the heterogeneity of C-agreement

Under the analysis we have developed here, the heterogeneity of C-agreement, far from being problematic, naturally falls out of a conspiracy of independently motivated structural properties which can vary both across languages & clause-types.

The directionality of probing falls out solely as a function of the relative position of the probe with respect to the CP phase. A ϕ -probe below the CP phase boundary will just probe downward (yielding DCA), while one above it will try and fail to probe downward, and then end up probing upward (yielding UCA/AA). Further variation across languages and clauses (may) depend(s) on the following factors. First, it is, at least in principle, possible (though perhaps not desirable) that the height of the CP phase varies across languages: this is ultimately an empirical question. Second, there is variation in the availability of particular ϕ -probes on Fin, Force and SAP and ϕ -goals like the addressee in SAP, which is conditioned by: (i) variation in the functional structure of the CP (which, in turn, falls out of selectional differences in the size of the clause embedded under distinct propositional predicates); (ii) whether a given C head is capable of hosting a ϕ -probe or not. Taken together, this suggests that the distribution of ϕ -probes and -goals in the C-domain should vary not only across individual clauses selected under distinct predicate-classes, but also across languages. For instance, a language like West Flemish must project a Fin head below the CP phase which can additionally host a ϕ -probe, yielding DCA. A language like Lubukusu, on the other hand, must project a more articulated CP which includes both Force and Fin, with the φ-probe being hosted on Force, above the CP phase, yielding UCA, but not on Fin. In languages like Basque or Tamil, embedded root CPs must project a High-C which hosts a ϕ -probe as well as a SAP which hosts the representation of the addressee as a ϕ -goal. Finally, in a language like English, we assume that C heads like Fin, Force and High-C are simply incapable of hosting ϕ -probes. Thus, regardless of the (non-)availability of particular C heads, we never get any variety of C-agreement.

6. Empirical predictions

Our proposal also makes several empirical predictions, which seem to be largely confirmed. Prediction A: Since both Fin and High-C can, in theory, host a ϕ -probe, a given language may instantiate both DCA and AA, albeit in mutually exclusive (embedded vs. root) clauses. This is confirmed for Upper Austrian German (Wiltschko and Heim 2016, Wiltschko 2014):

- (18) Wonn-ts nua es kumm-ts. (DCA) if-2PL only you.PL come-2PL 'If only you guys would come.'
- (19) Ea hot an neichn Hund, goi-ts. (AA)
 He has a new dog, CONF-2PL.ALLOC
 'He has a new dog, right (you guys)?'

Prediction B: since both Fin and Force may host ϕ -probes, DCA and UCA can co-occur in a single language or even in a single clause. Switch-reference systems might be an interesting instantiation of this prediction. For instance, switch-reference (SR) in Washo marks on C whether the matrix & embedded subjects are referentially distinct (20) or identical (21):

(20) Different subject (DS) $(-\check{s})$:

$$pro_i$$
 [Adele_j díme? sú:bi?-i- \S -ge] di-hámup'ay-é:s-i.
 pro_i Adele_j water 3/3.bring-IND-**DS**-NM.ACC 1/3-forget-NEG-IND 'I_i remember that Adele_j brought the water.'

(21) Same subject (SS) $(-\emptyset)$:

Adele_i [
$$pro_i$$
] dalá?ak ?-/'i:gi-yi- \emptyset -ge] hámup'ay-/'e:s-i. Adele_i [pro_i] mountain 3/3-see-IND-**ss**-NM.ACC] 3/3.forget-NEG-IND 'Adele_i remembers that she_i saw the mountain.'

Arregi and Hanink (2018, To Appear) argue that this instantiates C-agreement with both the embedded (DCA) & matrix (UCA) subjects, (see also Clem 2019:for a C-agreement treatment of SR in Amahuaca). Under the current proposal, such patterns could be derived by assuming that the embedded CP involves an articulated functional sequence with both Fin and Force, both of which are capable of hosting ϕ -probes. ϕ -probing downward from Fin would yield DCA while second-cycle upward ϕ -probing from Force would yield UCA. Additional (post-)syntactic mechanisms must then be adduced to ensure that these get spelled out as a single agreement morpheme (e.g. - δ or - \emptyset) on one of the C heads. Prediction C: UCA & AA can both occur in a given language, but should (all else being equal) be in complementary distribution in a given structure. If the ϕ -probe on Force appears in an embedded root clause, it should Agree with the addressee which will always be closer than a matrix argument. Thus, all else being equal, AA should bleed UCA. Indeed, we are aware of no counter-examples to this.²

References

Ackema, Peter, and Ad Neeleman. 2004. *Beyond morphology: Interface conditions on word formation*. Oxford: Oxford University Press.

van Alem, Astrid. 2020. Complementizer agreement in West-Germanic is clitic doubling. Ms., Leiden University.

²If, however, we still find what looks like UCA in such a configuration, it could be because: (a) the addressee in SAP is featurally invisible to the C-probe; (b) the UCA in question doesn't instantiate real agreement, or; (c) root clauses in the language never project as high as SAP. These alternatives should leave distinct empirical reflexes and are thus eminently testable.

Unifying C-agreement

- Antonov, Anton. 2015. Verbal allocutivity in a crosslinguistic perspective. *Linguistic Typology* 19:55–85.
- Arregi, Karlos, and Emily Hanink. 2018. Switch reference in Washo as multiple subject agreement. In *Proceedings of NELS 48*, 39–48.
- Arregi, Karlos, and Emily Hanink. To Appear. Switch reference as index agreement. *Natural Language and Linguistic Theory*.
- Baker, Mark. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.
- Baker, Mark. 2018. Toward a shared syntax for shifted indexicals and logophoric pronouns. *McGill Working Papers in Linguistics* 25:11–21.
- Bayer, Josef. 1984. COMP in Bavarian syntax. The Linguistic Review 3:209-274.
- Béjar, Susana, and Milan Rezac. 2009. Cyclic Agree. Linguistic Inquiry 40:35-73.
- Carstens, Vicki. 2016. Delayed valuation: A reanalysis of goal features, "upward" complementizer agreement, and the mechanics of case. *Syntax* 19:1–42. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/synt.12116.
- Carstens, Vicky. 2003. Rethinking complementizer agreement: Agree with a Case-checked goal. *Linguistic Inquiry* 34:393–412.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. by Michael Kenstowicz. Cambridge, Mass.: MIT Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford Studies in Comparative Syntax. OUP.
- Clem, Emily. 2019. Cyclic expansion in Agree: Maximal projections as probes. Ms. UCSD. Diercks, Michael. 2013. Indirect Agree in Lubukusu complementizer agreement. *Natural Language and Linguistic Theory* 31:357–407.
- Diercks, Michael, Marjo van Koppen, and Michael Putnam. 2020. Agree probes down: Anaphoric feature valuation and phase reference. In *Agree to Agree: Agreement in the Minimalist Programme*, ed. by Peter W. Smith, Johannes Mursell, and Katharina Hartmann, 347–389. Berlin: Language Science Press. DOI:10.5281/zenodo.354176.
- Driemel, Imke, and Maria Kouneli. 2020. C-Agree is local subject-verb agreement in Kipsigis. Ms., Universität Leipzig.
- Fuß, Eric. 2008. Multiple agreement and the representation of inflection in the C-domain. *Linguistische Berichte* 213:77–106.
- de Haan, Germen. 2001. More is going on upstairs than downstairs: embedded root phenomena in West Frisian. *Journal of Comparative Germanic Linguistics* 4:3–38.
- Haegeman, Liliane. 1992. *Theory and description in generative syntax*. Cambridge: CUP. Hill, Virgina. 2007. Vocatives and the pragmatics-syntax interface. *Lingua* 117:2077–2105.
- van Koppen, Marjo. 2005. One probe, two goals: Aspects of agreement in Dutch dialects. Doctoral dissertation, Leiden University.
- van Koppen, Marjo. 2017. Complementizer agreement. In *The Wiley-Blackwell Companion to Syntax Second Edition*, ed. by Martin Everaert and Henk van Riemsdijk, 923–962. London: Wiley-Blackwell.
- Krifka, Manfred. 2017. Assertions and judgements, epistemics, and evidentials. Workshop 'Speech Acts', Leibniz-ZAS.

- McFadden, Thomas. 2020. The morphosyntax of allocutive agreement in Tamil. In *Agree to Agree: Agreement in the Minimalist Programme*, ed. by Peter W. Smith, Johannes Mursell, and Katharina Hartmann, 391–424. Berlin: Language Science Press.
- Miyagawa, Shigeru. 2012. Agreements that occur mainly in main clauses. In *Main clause phenomena: New horizons*, ed. by Lobke Albrecht, Liliane Haegeman, and Rachel Nye, 79–112. Amsterdam: John Benjamins.
- Miyagawa, Shigeru. 2017. Agreement Beyond Phi. Cambridge, Mass.: MIT Press.
- Nformi, Jude. 2018. Complementizer agreement and intervention effects. Ms., University of Leipzig.
- Oyharçabal, Beñat. 1993. Verb agreement with non arguments: On allocutive agreement. In *Generative studies in basque linguistics*, ed. by José Ignacio Hualde and Jon Ortiz de Urbina. Amsterdam: Benjamins.
- Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20:365–424.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. by Liliane Haegeman, 281–337. Dordrecht: Kluwer Academic Publishers.
- Speas, Margaret. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 114:255–276.
- Speas, Peggy, and Carol Tenny. 2003. Configurational properties of point of view roles. In *Asymmetry in grammar*, ed. by Anna Maria DiSciullo, 315–344. Amsterdam: John Benjamins.
- Sundaresan, Sandhya. 2012. Context and (co)reference in the syntax and its interfaces. Doctoral dissertation, University of Tromsø (CASTL)/Universität Stutgart.
- Weisser, Philipp. 2019. Telling allomorphy from agreement. *Glossa: a journal of general linguistics* 4(1):86. 1–23.
- Wiltschko, Martina. 2014. *The universal structure of categories: towards a formal typology*. Cambridge, UK: Cambridge University Press.
- Wiltschko, Martina, and Johannes Heim. 2016. The syntax of confirmationals. In *Outside the clause: Form and function of extra-clausal constituents*, ed. by Gunther Kaltenböck, Evelien Keizer, and Arne Lohmann, 305–340. John Benjamins.
- Zeijlstra, Hedde. 2012. There is only one way to Agree. *The Linguistic Review* 29:491–539. Zwart, C. Jan-Wouter. 1997. *Morphosyntax of verb movement: A Minimalist approach to the syntax of Dutch*. Dordrecht: Kluwer.

Thomas McFadden, Sandhya Sundaresan mcfadden@leibniz-zas.de, sandhya@sndrsn.org