Rara and theory testing in typology

The natural evolution of non-canonical agreement

This is a pre-print version of the paper published as:

Round, Erich R. 2019. “Rara and Theory Testing in Typology: The Natural Evolution of Non-Canonical Agreement.” In *Morphological Perspectives: Papers in Honour of Greville G. Corbett*, edited by Matthew Baerman, Oliver Bond, and Andrew Hippisley, 414–46. Edinburgh: Edinburgh University Press.

In typology, *rara* provide valuable tests for theoretical hypotheses. Here I consider the *rarum* of **PERSON** inflection in Kayardild, which has only two surface contrasts but is found across all words in complementized subordinate clauses. I introduce a general schema for reasoning about the diachronic emergence of *rara*, and reconstruct the evolution of Kayardild subordinate **PERSON** agreement, from an earlier state in which a main-clause inverse system was coupled to a system of complementizing **CASE** agreement. Serendipitously, the same synchronic facts have been analysed twice earlier without the benefit of the full diachronic backstory, and so present a retrospective case study in what diachrony offers for the analysis of *rara*, structures which by definition are difficult to contextualize using synchronic typology alone. I argue that since *rara* are so valuable for the testing of typological theories, and since diachrony may offer the only source of convincing explanation for them, it follows that typological science will need to refer to diachrony for the successful development of theory. It cannot rely on synchrony alone.¹

Keywords: *rara*, person agreement, inverse systems, morphological change, subordinate clauses, Kayardild, Lardil, Yukulta

1 Generalization, theory and *rara*

Typological research seeks to make valid generalizations about the diversity of human language. It proceeds from observed linguistic traits, their distributions across geography and genealogy, and their relationships to other traits. Typological theories generate hypotheses, whose testing can prompt us to make more precise and systematic observations. Theories also propose explanations

¹A version of this paper was presented at the New Fields for Morphology Workshop, Melbourne 2016. My thanks to the audience there and to Nick Evans for discussion that has improved the paper. All shortcomings are my own. This research was supported by Australian Research Council grant DE150101024 to Erich Round.

for observations. Corbett (2012) for example generalizes over extensive typological evidence to propose a theory of features, according to which, inflectional features are limited to only a handful of kinds. As in all sciences, theories in typology have apparent counter-examples. Rare linguistic traits, or *rara* (Plank 2000; Cysouw and Wohlgemuth 2010), often present such instances. If so, they challenge the theory, and warrant closer examination. Here I present an instance in which Corbett's theory prompted the re-examination of a *rarum*, namely the inflectional marking of complementized clauses in Kayardild. This led to a revised understanding of the phenomenon, and to support for the theory. Moreover, the re-examination presents the opportunity to compare three analyses of the same facts. The comparison highlights the methodological value for typology of attending to diachrony when evaluating *rara*. The argument has been made before, that *rara* often arise through normal processes of historical change, only that because they require unlikely combinations of multiple changes, or rare preconditions, they are less likely to arise than common traits (Harris 2008; Harris 2010; Grossman 2016). Here I emphasise two corollaries of the nature and scientific value of *rara*. Because *rara* are particularly useful for hypothesis testing and attendant theoretical progress, and since our understanding of *rara* often requires reference to diachrony, it follows that methodologically, typology should ideally include as much about diachrony as it can in its evidentiary base; and epistemologically, it would be mistaken to characterize typology as an ahistorical discipline (e.g. Daniel 2011), since diachrony is particularly important for understanding some of the best evidence for testing typological theories.

The *rarum* at the centre of the discussion will be 'subordinate clause complementization marking', or SCCM, in Kayardild. In Kayardild, certain subordinate clauses (SCs) are morphologically marked as complementized (Evans 1995a; Round 2013). The complementized–uncomplementized opposition conveys information about co-reference between arguments of main and subordinate clauses, and about topic continuity in discourse (Evans 1995a). As it happens, in Kayardild there are two ways to mark a SC as complementized, by using two different inflectional markers. Both of the markers descend historically from exponents of **CASE**, and when they are used to mark complementization they attach to every word, or nearly every word, in the complementized SC (in the case of subject pronouns, the pronoun is selected from an appropriate series; see the Appendix for conditions under which words escape marking for SCCM). The choice between these two types of SCCM is decided by the **PERSON** value (including clusivity) of the SC subject, and, if the subject is second person, according to whether 'the speaker wants to group him/herself with the addressee' (Evans 1995a:494), a factor which I will refer to as **SOLIDARITY**. The first type of marking is illustrated for a first exclusive subject in (1), third person subject in (2), and second person subject (non-solidarity) in (3). The second type is illustrated for a second person subject (solidarity) in (4). For the moment, the markers of complementization are glossed COMP1 and COMP2. (Examples are given as: practical orthography on line one; underlying phonological forms on line two; then comparative morphological categories, discussed further in §3; and the function of morphs on the final line.)²

²Notes regarding interlinear glossing: [1] The phonological line distinguishes two types of morpho-phonological boundary, written '-' and '+'. Different phonological processes apply depending on the boundary type (Round 2009; Round 2013:12–14). [2] Not all functions have corresponding overt morphological exponents (cf. **SINGULAR NUMBER** in English nouns). Such functions are placed in square brackets on the fourth line. [3] Some functions correspond to multiple pieces of morphology. These one-to-many relationships are glossed in braces on the third and fourth lines. [4] The 'termination' element, glossed T on the third line of examples, is a piece of morphology with no meaning, and so is not glossed on line four. [5] Abbreviations follow the Leipzig Glossing Conventions; additional abbreviations are: ACT actuality mood, COLL collative case, COMP complementization, DON donative case, INVO object of inverse construction, K. Kayardild, L.

Person/solidarity	1excl	1incl	2(sol.)	2(non-sol.)	3
SCCM type	1	2	2	1	1

Table 1: Choice of complementization marking type in Kayardild

- (1) Thardabadija wuranki, [ngarrwaa ngakathurrk].
K. ṭaṭa-patic-a wuṭan-ki ṇa-r-wa-a ṇakaṭ-kurka
shoulder-carry-UNM food-XERGLOC 1-DU-XOBL-T catch-XERGLOC:XDAT
shoulder-carry-UNM food-TRO 1-DU-COMP1[SBJ] catch-PRES:COMP1
‘We’re carrying the food on our shoulders, that we’re catching.’ [R2005-jul14a]
- (2) Balmbu nyingka kurriju, [ngaakantha dangkantha
K. palmpu ṇiṇ-ka kuric+ku ṇa:ka-ṇṭa ṭaṇka-ṇṭa
tomorrow.XPROP 2sg-T see-XPROP which-XDAT man-XDAT
tomorrow.FUT 2sg.SBJ see-FUT which-COMP1[SBJ] man-COMP1[SBJ]
- ngijinjinaantha wangalwujarranth.]
ṇicip+kina:-ṇṭa waṇalk-wuc-ṇara-ṇṭa
1SG.XOBL-XABL-XDAT boomerang-DON-XPST-XDAT
1SG-PST-COMP1[TRO] boomerang-DON-PST-COMP1
‘Tomorrow I will show you the man who gave me this boomerang.’ [W1960]
- (3) Jinaa bijarrb, [ngumbaa kurulutharranth] ?
K. cina-a picarpa ṇuṇ+pa-a kuṭuluṭ+ṇara-iṇṭa
where-T dugong 2SG-XOBL-T kill-XPST-XDAT
where dugong 2SG-COMP1 kill-PST-COMP1
‘Where is the dugong which you killed?’ [E493.ex.12-13b]
- (4) Jinaa bijarrb, [nyingka kurulutharray] ?
K. cina-a picarpa ṇiṇ+ka kuṭuluṭ+ṇara+ki-a
where-T dugong 2SG-T kill-XPST-XERGLOC-T
where dugong 2SG[COMP2] kill-PST-COMP2
‘Where is the dugong which you killed?’ [E493.ex.12-13a]

The correlation between **PERSON** and **SOLIDARITY** values, and the two SCCM types is summarized in Table 1.

This chapter will be the fourth work to consider the analysis of SCCM in Kayardild. The proposal here is that SCCM is **PERSON** agreement, with some conditioning by **SOLIDARITY**. Supposing we accept this analysis, it can be noted that **PERSON** agreement in Kayardild is distinctly non-canonical, in the sense of Canonical Typology (Corbett 2005; Brown, Chumakina, and Corbett 2013). It marks **PERSON** in a

Lardil, NTR non-transitive, NTRS non-transitive subject, T termination, TRO transitive object, UNM unmarked tense, XABL ablative comparative morphomic category, XDAT dative comparative morphomic category, XERGLOC ergetive–locative comparative morphomic category, XGEN genitive comparative morphomic category, XNOMZ nominalization comparative morphomic category, XOBL oblique comparative morphomic category, XPAST past comparative morphomic category, XPROP proprietive, Y. Yukulta. [6] Example sources are abbreviated as: [W1960] for Wurm’s (1960) Kayardild recordings, [E+page; example] for Evans’ (1995a) Kayardild material, [R+tape date & number] for Round’s (2005; 2007) Kayardild recordings, [K+tape number] Keen’s (1970) Yukulta recordings, my transcriptions.

non-canonical location (across the whole clause, and only in complementized SCs), with a curious marker (etymological **CASE**) and with a non-canonical set of syncretisms that results in only a binary surface opposition (type 1 vs 2) for an underlying, four-valued distinction (1INCL, 1EXCL, 2, 3). Much of the chapter will be concerned with explaining how, just as for many *rara*, this non-canonical outcome arose out of natural historical changes. Interestingly, this is also the first occasion on which it is proposed that SCCM in Kayardild is **PERSON** agreement. In previous analyses, Evans (1995a; 2003) analyses SCCM as **CASE** agreement, where the choice between the two **CASE** values (corresponding to SCCM types 1 and 2) is conditioned by **PERSON** and **SOLIDARITY**, though Evans (2003) points out that it is unclear what the controller of agreement would be, or alternatively if SCCM is a matter of government, what the governor would be. Round (2013) accounts for SCCM with two inflectional features, **COMPLEMENTIZATION** and **SEJUNCT**, the former associated syntactico-semantically with all complementized SCs and the latter with first exclusive, second (non-solidarity) and third person subjects. These analyses are compared towards the end of the chapter.

The chapter is arranged as follows. §2 sets out a framework to be used in later sections for discussing the historical emergence of non-canonical linguistic patterns through natural diachronic changes. §3 introduces the Tangkic language family and §4 its main inflectional traits. §5 reconstructs the historical development of SCCM. §6 considers the emergence of SCCM in terms of its shifting synchronic systems of agreement. §7 reviews earlier analyses formulated in the absence of the historical backstory, and §8 concludes.

2 Non-canonical outcomes of natural historical changes

It has long been observed that unremarkable historical changes will at times give rise to synchronic rules, or patterns of alternation, that are typologically odd (Wang 1968; Bach and Harms 1972; Givón 1979; Anderson 1981; Anderson 1988). A common form that these changes take involves the loss of a mediating node in a causal chain. For example, prior to the change, an opposition between meanings A and A' is correlated with, or causes, an alternation between structures X and X', and moreover the alternation between X and X' causes a secondary alternation between structures Y and Y'. This is shown schematically in Figure 1a.

Figure 1: Change in causal links

For instance, in the pre-Old English nominative of *fōt* ‘foot’, a semantic alternation in **NUMBER** between SINGULAR (A) and PLURAL (A') was correlated with, or caused, an alternation in overt inflection between no suffix (X) and a suffix *-i* (X') which in turn caused a phonological alternation between a root vowel *ō* (Y) and *ø* (Y'). At this point, the causal link between {SG ~ PL} and {-∅ ~ -i} was one of relatively canonical inflection (Corbett 2007) and the causal link between suffixal {∅ ~ -i} and root vowel {*ō* ~ *ø*} was a matter of common vowel assimilation. Each of the causal links in Figure 1a is typologically unremarkable.

From this initial state, a change occurs. If that change obliterates the opposition between X and X',

such as when the pre-Old English PLURAL suffix *-i* was lost, then the causally mediating node $\{X \sim X'\}$ is removed from the network. A possible response by subsequent generations is to create a new causal link directly from $\{A \sim A'\}$ to $\{Y \sim Y'\}$, as in Figure 1b. For Old English *fōt*- ‘foot’, this meant that the $\{\text{SG} \sim \text{PL}\}$ alternation now directly caused root vowel apophony $\{\bar{o} \sim \bar{\theta}\}$. The new causal link was non-canonical: inflection was now realized in a non-canonical location (root medially) and by non-canonical means (apophony rather than affixation). A cognitive constraint upon this historical outcome is that speakers must be able to draw a new causal link from $\{A \sim A'\}$ directly to $\{Y \sim Y'\}$.

Variants on the basic motif are imaginable. The initial situation may involve a more complex cause which gets simplified as in Figure 2a,b; or the co-opting and establishment of a new cause (Figure 2c,d); or a partial mix of these (Figure 2e,f). In all cases, the constraining factor will be whether a new generation of speakers is able to infer a new causal network (Figure 2b,d,f), given only the evidence handed to it by history.

Figure 2: More complex changes in causal links

Non-canonicity can also arise through restructuring of fine-grained mapping relationships within a single causal link. For example, suppose that prior to a change, the alternation in meaning is between four values $\{A \sim A' \sim A'' \sim A'''\}$ and the original mediating node likewise has four alternants $\{X \sim X' \sim X'' \sim X'''\}$. The link from A values to X alternants is canonical, one-to-one. Suppose also that the secondary alternation was one of assimilation between two vowels heights, and so involved just two alternants, $\{Y \sim Y'\}$. All of the causal links and mappings in this pre-change system are canonical. However, if the mediating node is lost and the four values $\{A \sim A' \sim A'' \sim A'''\}$ are then mapped directly onto just two alternants $Y \sim Y'$, the mapping now has a less canonical, many-to-one geometry.

In grammars causal links take many forms: the assignment of grammatical roles based on (i.e., caused by) semantic roles; morphological case based on grammatical roles; inflectional features of an agreement target based on features of an agreement controller; phonological exponence based on an in-

flectional feature structure; or the surface form of a phonological target based on its phonological environment. Important here is that firstly, any of these causal relationships may fill the role of ‘causal link’ in the schemata above, and secondly, that the process of historical restructuring may involve older causal links of one type being supplanted by new links of a qualitatively different type (Anderson 1992). Some types of qualitative shifts have established labels such as ‘phonologization’, or ‘morphologization’, but the range of conceivable types extends beyond those with established linguistic terminology.

In the following sections we will see in the prehistory of Kayardild the loss of a mediating causal node and the restructuring of the causal links in the SCCM system.

3 Tangkic

The Tangkic family of languages traditionally was spoken in the southern Gulf of Carpentaria in northern Australia, on the small Wellesley Island group and the adjacent, low-lying mainland. The description in §3 is based closely on Round (2017).

Figure 3: Tangkic family tree after Evans (2005)

The customary family tree for Tangkic (Evans 2005) is in Figure 3. Relatively well described are Lardil (Hale 1973; Klokeid 1976; Ngakulmungan Kangka Leman 1996), Kayardild (Evans 1992; Evans 1995a; Round 2009; Round 2013; Evans and Round 2017) and Yukulta (a.k.a Gangalidda; Keen 1983; Nancarrow 2014; Round 2014). Less well studied are Yangkaal (a.k.a. Yangkaralda), for which some documentation exists (Cook 2017; Cook and Round 2017), and Yangarella and Nguburindi for which only early colonial wordlists are available. Round (2017) is a comparative overview of Tangkic phonology, morphology and morphosyntax.

Historical relationships among the Tangkic languages are a matter of ongoing research. While the tree in Figure 3 is a reasonable representation of relatedness among the Tangkic languages in terms of lexicon, the history of Tangkic morphosyntax is complex. Round (2017) highlights strong resemblances between Kayardild (and Yangkaal) and Lardil in distinction from Yukulta, contradicting the relationships implied by Figure 3. Memmott et al. (2016) suggest, based also on phylogenetic analysis of vocabulary, that Tangkic may possess an ‘Eastern’ clade that resulted from a significant prehistoric

contact event, resulting in a phylogenetic network as in Figure 4. Kayardild may well have inherited its vocabulary via the lower dashed arrow in Figure 4, but much of its morphosyntax via the upper dashed arrow.

Figure 4: Tangkic family tree after Memmott et al. (2016)

Returning to the synchronic plane, as in many of Australia’s languages (Nordlinger 2014), much of the syntactic structure in Tangkic sentences is reflected not in word order, but in inflectional morphology. Notably, inflectional features often take scope over a syntactic constituent such as DP, VP or even a whole clause, within which all words inflect overtly for that feature. For example, in (5) the final two words represent one DP embedded in another.

- (5) Ngada latha karnjini [[marunnganku] maarnku]
L. ɲata ɭaɬ-a kaŋcin-iŋɬa maɭun-ŋan+kuɭu ma:ŋ+kuɭu
 1SG spear-UNM wallaby-XDAT boy-XGEN-XPROP spear-XPROP
 1SG.SBJ spear-UNM wallaby-TRO boy-GEN-INS spear-INS
 ‘I speared the wallaby with the boy’s spear.’ [Richards (2013) ex.3]

The word *marunnganku* in the lower DP inflects for GENITIVE **CASE** (glossed on the lowest line as GEN). All words in the matrix DP, including *marunnganku* in the lower DP, inflect for INSTRUMENTAL **CASE** (glossed on the lowest line as INS). Consequently, the Tangkic languages are well-known for their *Suffixaufnahme* (Evans 1995b; Richards 2013; Round 2013), the ability of a single word to accrue inflectional markings for multiple features, associated with multiple syntactic nodes that dominate it.

All Tangkic languages allow words in DPs to inflect for **CASE**. Core argument alignment is nominative–accusative in Lardil, Kayardild and Yangkaal. The more complex, mixed system of Yukulta is discussed in §4.4. A system of complementization marking in subordinate clauses (SCs) functions to signal co-reference relationships between SCs and their main clauses (MCs). In Kayardild, SC complementization also serves discourse functions; possible discourse functions of complementization in other Tangkic languages are poorly understood, and in this chapter I focus primarily on the syntactic use of complementization, to signal co-reference in MCs and SCs. In all Tangkic languages, **TENSE** in SCs is marked not only on the verb but on the majority of non-subject DPs. In Lardil, Kayardild and Yangkaal, this is also true for **TENSE** in MCs (Klokeid 1976; Evans 1995a; Nordlinger and Sadler 2004; Round 2013; Round and Corbett 2017).

		Underlying allomorph, after		
		C	Front V	Back V
Lardil	acc case, comp	-iṇṭa	-iṇṭa	-iṇṭa
Kayardild	obl case, several tenses, comp	-iṇca	-ṇca	-ṇṭa
Yukulta	dat case, comp	-iṇca	-ṇca	-ṇṭa

Table 2: Sets of forms and functions of the comparative category xDAT

Tangkic word forms can be complex. Fully inflected words consist of a lexical stem followed by zero or more inflectional suffixes, and potentially end with a meaningless, monosyllabic ‘termination’ element (Round 2013; Round 2017). Because Tangkic words often contain multiple layers of inflection, it is useful to distinguish morphologically between lexical stems (prior to any inflection) and inflected stems, which themselves may serve as bases for more inflection. All Tangkic stems, whether lexical or inflected, can be classed either as thematic or as athematic. Inflectional allomorphy is often sensitive to the distinction. In (6) for example, a Yukulta SC (shown in square brackets) is inflected for **PAST TENSE**. The exponent of **PAST TENSE** is /+ṇarpa/ on the thematic stem /kurkaṭ-/ but /+kinapa/ on the athematic stem /mija|/.

- (6) Gurrija =nganda [miyarlinaba gurrgharrba].
 Y. kuric-a =ṇa-nta mija|+kinapa kurkaṭ+ṇarpa
 see-UNM spear-XABL take-XPST
 see-UNM =1SG.TRS>2SG.TRO-PST spear-PST take-PST
 ‘I saw you take the spear’ [K05-001917A]

The inflectional systems of Tangkic languages are striking for their use and re-use of the same suffixal forms, and sets of form alternants, for multiple functions. Round (2011; 2013; 2015; 2016) argues based on Kayardild, that this systematic tethering of sets of forms to entire sets of functions is ‘morphomic’ in the sense of Aronoff (1994).³ For expository purposes here, it will be useful to refer to what Hartwig (2014) has called ‘comparative morphomic categories’ in Tangkic. Each category comprises a set of forms. Those sets can be identified as cognate across the Tangkic family, and each set will have some number of functions in each modern language. Following Hartwig, I assign these form-set categories a label beginning with X, for example xDAT for the set of forms in Table 2.

Comparative morphomic categories have multiple forms and multiple functions in each language, so it is impossible for labels like ‘xDAT’ to convey an exhaustive characterisation of the categories’ functions. Nevertheless, a basic mnemonic system is used, so that ‘xDAT’ for example is the label for the form-set, one of whose functions in proto Tangkic was to mark **DATIVE CASE**. Similarly, ‘XERGLOC’ labels the form-set, one of whose functions in proto Tangkic was to mark the homophonous **ERGATIVE** and **LOCATIVE CASE**. Even in proto Tangkic though, the form-sets would have had multiple functions, often including the marking of **CASE**, **SC TENSE** and **SC complementization**. Comparative morphomic categories appear in the third line of example sentence glosses.

³More narrowly, they are ‘meromorphomes’: morphomic morphological pieces from which word forms are constructed, in contrast to ‘rhizomorphomes’ which are morphomic classes of roots, and ‘metamorphomes’ which are morphomic sets of cells within a paradigm (Round 2015).

4 Clausal morphosyntax in Tangkic

4.1 Overview

The aim of §4 is to acquaint reader with the major aspects of inflectional morphology in Tangkic main and subordinate, declarative clauses (for a more comprehensive overview see Round 2017). The focus is on aspects of inflection which will be instrumental for reconstructing the history of subordinate clause complementization marking, SCCM. Though there are many details to introduce here, I place special emphasis on those facts that are most relevant to the historical reconstruction; they are the main points to absorb from each subsection. Any successful analysis of Tangkic clauses needs to distinguish carefully between **CASE** categories (such as **ERGATIVE**), versus the marking of such categories (such as by **XERGLOC**), versus the more underlying grammatical roles which map onto such categories (such as the ‘transitive subject’ role). In the analyses presented below, I assume that semantic predicates have semantic arguments, such as the more actor-like and less actor-like arguments of the two-place predicate ‘hit’. These semantic arguments map onto to certain grammatical roles. In Yukulta, a key distinction is between the Transitive Subject role, abbreviated and glossed as **TRS**; the Transitive Object role (**TRO**); and the Nontransitive Subject role (**NTRS**). Lardil and Kayardild typically treat **TRS** and **NTRS** equivalently, as a unitary Subject role **SBJ**. In addition, goal-like and recipient-like arguments of three-place predicates typically map to a grammatical Goal role (**GOAL**). Grammatical roles are posited because they interact with morphology and syntax in distinctive ways. The mapping from grammatical roles to morphological **CASE** categories can be affected by clause-level factors such as mood and polarity, and DP-level factors such as the distinction between pronominals and full nominal DPs. To facilitate comparison between the individual languages, the terminology I use is comparative. I will refer to the grammatical roles just introduced, and the comparative morphomic categories such as **XDAT**, rather than referring to the various language-specific labels that have been employed in the literature on individual languages.

The remainder of §4 provides a gradual introduction to the morphosyntax of the language family. §4.2 begins with the shared core of subordinate clause morphosyntax; §4.3 moves to the etymologically related, main clauses of Lardil and Kayardild; §4.4 shifts to the historically more conservative and more complex main clauses of Yukulta.

4.2 Subordinate clauses with no SCCM

Subordinate clauses (SCs) are similar across the Tangkic language family. All languages distinguish complementized SCs, with SCCM, from uncomplementized SCs, without SCCM. We return to complementized SCs in §5.

In all Tangkic SCs, including those without SCCM, subjects are morphologically unmarked for **CASE** (if pronominal, they are selected from the basic-stem series of pronouns) and objects are inflected not for **CASE** but for **TENSE**. The three main **TENSE** categories, typically interpreted relative to the tense of the main clause (MC), are **SIMULTANEOUS**, **PAST** and **FUTURE**.⁴ Subordinate **SIMULTANEOUS TENSE** is

⁴Other subordinate clauses include apprehensives, which express events to be feared or avoided (cf. Dixon 1980:380, Blake 1987:136); desideratives; and in the Southern Tangkic languages, purpose-of-movement clauses. The basic tenses are used also with certain extended senses, such as **PAST** for conditional protases, and **FUTURE** for desires, potentials and commands.

marked by XERGLOC; FUTURE by XPROP; and PAST by XPST in Lardil, or in Yukulta and Kayardild by XPST on thematic stems and XABL on athematics. Examples below illustrate the SIMULTANEOUS in Yukulta (7), PAST in Lardil (8) (see also (6), above, in Yukulta), and the FUTURE in Kayardild (9).⁵

- (7) Ngada =nganbungarri, gurri [gamburiji.]
 Y. ŋaɬ-ta =ŋanpuŋa-ri kuri kampu.ɟic+ki
 1SG-T watch talk-XERGLOC
 1SG[TRS] =1SG.TRS>3PL.TRO-TR.PRES watch[UNM] talk-SIMUL
 ‘I’m watching them talking. [K5-001918A]’

- (8) Mangarda yuurr-denja [wangalkarr wungitharr.]
 L. maŋ-ka=ɬaŋka jur-ɬæɲcac waŋalk+ŋarpa wuɲic+ŋarpa
 child-T=person PRF-run boomerang-XPST steal-XPST
 child[SBJ] PRF-run[UNM] boomerang-PST[TRO] steal-PST
 ‘The child who stole a boomerang has run off.’ [Klokeid 1976:142]

- (9) Nyingka karngija, [marndiinangku,]
 K. ɲiŋ-ka karɲic-a maŋɬi<i>c-ŋaŋ+ku:
 2SG-T hold-UNM deprive<PASS>-XNEG-XPROP
 2SG.SBJ hold-IMP deprive-PASS-NEG-FUT
 ‘You look after (your country), so you won’t be robbed of it. [Evans 1995:523]

4.3 Main clauses in Lardil and Kayardild

Lardil and Kayardild also have **TENSE**-marked objects in MCs. Evans (1995) argues that these **TENSE**-marked MCs emerged historically through the grammaticalization of erstwhile uncomplementized SCs (§4.2), which increasingly took on ‘insubordinated’ MC functions. MCs inflected for FUTURE and PAST **TENSE** are shown in (10,11).

- (10) Ngada werethu wangalku.
 L. ŋata wæ.ɬæc+ku.ɬu waŋalk+ku.ɬu
 1SG throw-XPROP boomerang-XPROP
 1SG.SBJ throw-FUT boomerang-FUT[TRO]
 ‘I will throw the boomerang.’ [Hale 1996:23]
- (11) Bilda kabatharra wurankina kuruna?
 K. pi-l-ta kapaɬ+ŋara wu.ɬan+kina ku.ɬu+kina
 3-PL-T find-XPST food-XABL egg-XABL
 3-PL.SBJ find-PST food-PST[TRO] egg-PST[TRO]
 ‘Have they found any eggs?’ [Wurm 1960]

Historically speaking, the FUTURE and PAST **TENSE** MCs of Lardil and Kayardild are direct inflectional continuations of old SCs. The same cannot be said of PRESENT **TENSE** MCs. In Lardil, PRESENT **TENSE**⁶ MCs have objects inflected with XDAT, and verbs with an ‘unmarked’ suffix as in (12).

⁵In Lardil and Kayardild, the SIMULTANEOUS SCs differ from those in Yukulta, sharing their inflectional properties with MC PRESENT **TENSE** clauses, described in §4.3.

⁶I use the term ‘present tense’ for comparative purposes. In modern Lardil, this clause type is a default non-future.

(12)	Diin	ngawa	betha=ku	ngithan	thakin	jayin.
L.	ti:n	ɲawu	bæc-a=kun	ɲiɬa-iɲɬa	ɬak-iɲɬa	ca-iɲɬa
	this	dog	bite-UNM=ACT	1SG.XOBL-XDAT	left-XDAT	foot-XDAT
	this[SBJ]	dog[SBJ]	bite-UNM=ACT	1SG-TRO	left-TRO	foot-TRO
	'This dog bit my left foot.' [NKL1996:212]					

As we will see in §4.4 next, this set of inflections (xDAT on objects, and ‘unmarked’ verbs) and is the last remaining hold-over in Lardil of a much older MC inflectional system, one which remains intact in its more conservative cousin, Yukulta.

PRESENT **TENSE** MCs in Kayardild are mixed. Objects are inflected with XERGLOC, cognate with SC SIMULTANEOUS **TENSE** marking in Yukulta. Curiously enough though, Kayardild’s PRESENT **TENSE** verbs, like those in Lardil, are ‘unmarked’, as in (13).⁷

(13)	Dathina	dangkaa	makuya	thardaya	buruth.
K.	ɬaɬina	ɬaɲka-a	maku+ki-a	ɬaɬa+ki-a	puɬuɬ-a
	that	man-T	woman-XLOC-T	shoulder-XLOC-T	grab-UNM
	that[SBJ]	man[SBJ]	woman-PRS[TRO]	shoulder-PRS[TRO]	grab-UNM
	'That man is grabbing the woman by the shoulder.' [W1960]				

At this stage our picture of Tangkic morphosyntax is already accruing a number of idiosyncrasies. The primary point to take away from §§4.2–4.3 is that all Tangkic SCs have objects inflected for **TENSE**; Lardil and Kayardild now also use these inflectional patterns in MCs, at least in the non-present tenses.

4.4 Main clauses in Yukulta

MCs in Yukulta are quite different to those of Lardil and Kayardild. **TENSE** is not marked on objects, and not even on verbs, but on a second-position clitic; and a major division separates transitive MCs (TRMCs) from ‘nontransitives’ (NTRMCs). The NTRMCs in turn can be divided into ‘regular’ NTRMCs and ‘inverse’ NTRMCs. For expository reasons, it is simplest to begin with the regular NTRMCs. As it happens, these are the etymological source of Lardil’s PRESENT **TENSE** marking (with xDAT on objects and unmarked verbs) which we saw in §4.3. Yukulta’s regular ‘nontransitive’ MCs (Round 2014) include monovalent intransitives as well as bivalent MCs under specific, classically low-transitivity semantic conditions (Hopper and Thompson 1980): in irrealis mood, in the context of certain negative verbal inflections, as a strategy for formulating polite desideratives, and in clauses whose verbal predicate is drawn from the small class meaning ‘wait for’, ‘fear’, ‘talk to’, ‘be sorry for’ and ‘dream or think of’ (Keen 1983). Subjects of NTRMCs have a NTRS grammatical role. NTRS full DPs have no overt **CASE** inflection (and pronouns take the basic-stem form). Sentence (14) shows an apposed, basic-stem pronoun *ngagulda* and full DP *burldamurra* in the NTRS role.

⁷ An inflectional equivalent of Yukulta’s SC SIMULTANEOUS does occur rarely in Kayardild MCs, denoting events happening right at the point of utterance.

(14)	Ngagulda	=gula	burldamurra	warraja.
Y.	ŋa-ku-l-ta	=kul-a	pul[amur-a	warac-a
	1-INCL-PL-T		three-T	go-UNM
	1-INCL-PL[NTRS]	=1INCL.PL.TRS-PRS.NTR	three[NTRS]	go-UNM
	'We three can go.' [K3-001789B]			

Objects of regular NTRMCs are in the GOAL grammatical role, and are assigned DATIVE CASE, marked by XDAT on full DPs and by a dative series of pronouns. In (15) nontransitivity is due to the presence of the lexical predicate *bulwiji*- 'fear', and in (16) is triggered by negative inflection on the verb. The nontransitive status of Yukulta's NTRMCs is also marked on the second-position clitic.

(15)	Bulwija	=rnala	gunawuna	dathininja.
Y.	pulwic-a	=ŋa-l-a	kunawuna	taŋin-iŋca
	fear-UNM		child	that-XDAT
	fear-UNM	=3SG.GOAL-3PL.NTRS-PRS.NTR	child[NTRS]	that-GOAL
	'The children are scared of that (man).' [K3-001789B]			

(16)	Warlirra	=ngga	dalmatharri	ngijinja.
Y.	waŋira	=ŋka	taŋmaŋ+wari	ŋic-iŋca
	NEG		chop-XPRIV	wood-XDAT
	NEG	3SG.NTRS.PRES[3SG.GOAL]	chop-NEG	wood-GOAL
	'He's not chopping wood.' [K5-001917B]			

In addition to its 'regular' NTRSCs, Yukulta has a second kind of nontransitive MC, whose use is required in bivalent clauses with 'inverse' combinations of subject and object **PERSON** values. The general triggering condition for inverse NTRSCs is that the object is higher than the subject on Yukulta's person hierarchy $1 > 2 > 3$. Inverse objects are assigned an Inverse Object (INVO) grammatical role, and because of the way the inverse is triggered, an inverse object will necessarily be either FIRST or SECOND **PERSON**.

Cross-linguistic evidence shows that inverse systems are often partly idiosyncratic (Jacques and Antonov 2014; Gildea and Zúñiga 2016; Witzlack-Makarevich et al. 2016), and Yukulta is no exception. In Yukulta, SECOND **PERSON** subjects acting specifically upon FIRST **PERSON** *singular* objects do not trigger the inverse.

In terms of inflectional marking, subjects in inverse NTRMCs appear as in any other NTRMC, while the marking of objects is what makes the inverse inflectionally distinctive. Inverse objects are realised morphologically by a pronominal stem inflected with the XERGLOC marker. Any apposed, full-DP objects are also marked with XERGLOC, as in (17), where the inverse object is 'us three'. Like other NTRMCs, inverse NTRMCs are marked as nontransitive in the second-position clitic.

(17)	Jinkaja	=guluwaninggi	ngaguluwanji	burldamurri.
Y.	cinkac-a	=kuluwa-ni-ŋki	ŋa-ku-lu+paŋ+ki	pul[amur+ki
	follow-UNM		1-INCL-PL-XOBL-XERGLOC	three-XERGLOC
	follow-UNM	=CLITIC	1-INCL-PL-OBL-INVO	three-INVO
	(CLITIC: 1INCL.PL.GOAL-3S.NTRS-FUT.NTR)			
	'He will follow us three.' [K9-001841B]			

Overall then, the inverse is distinguished by the marking of MC objects with XERGLOC instead of usual

possible MC arguments, is co-referential with some argument in the SC; the SC then agrees in **CASE** with that MC argument. For example, in (20) from the Panyjima language of Western Australia, the SC subject is co-referential with the MC object; the MC object has accusative **CASE**, which the entire SC then agrees with (Dench and Evans 1988:28–29).

(20)	Ngatha	wiya-rna	ngunha-yu	marlpa-yu
Panyjima	1SG.NOM	see-PST	that-ACC	man-ACC
	paka-lalha-ku	nharniwali-ku	warrungkamu-la-ku.	
	come-PERF-ACC	here.ALL-ACC	morning-LOC-ACC	
	'I saw that man who came this way this morning.'			

SCCM in the modern Tangkic languages is never entirely like complementizing **CASE** agreement in this sense, though it bears some striking similarities.

5.2 SCCM in Yukulta

In Yukulta, SCCM signals co-reference relationships between a SC subject and certain MC arguments. We saw in §4.4 that the marking of MC arguments in Yukulta is sensitive to MC transitivity. SCCM in Yukulta is also sensitive to transitivity in the matrix MC. When the matrix MC is transitive, overt SCCM is used when there is co-reference between the SC subject and the MC subject; the SCCM marker is XERGLOC, and it appears on every word of the SC. An example is shown in (21).

(21)	Dangkaya	=ganda	balatha	jardabu,
Y.	ṭaŋka+ki-a	=kanta	palaṭ-a	caṭapu
	person-XERGLOC-T		kill-UNM	emu
	person-ERG[TRS]	TR.PST[3SGS>3SGO]	kill-UNM	emu[TRO]
	[wurlankinabaya		diyajarrbaya.]	
	wuḷan+kinapa+ki-a		ṭiac+kinapa+ki-a	
	food-XABL-XERGLOC-T		eat-XPST-XERGLOC-T	
	food-PST-COMP[TRO]		eat-PST-COMP	
	'The man _i killed the emu after he _i 'd eaten the food' [K5-1918A]			

When the matrix MC is nontransitive, overt SCCM is used when there is co-reference between the SC subject and the MC object.⁸ The SCCM marker now is xDAT, which appears on every word of the SC. An example is in (22).

⁸There is also a default case, which is not directly relevant to the present discussion: if the SC subject is co-referential with neither the MC subject nor the MC object, then the SC is complementized by default and the SCCM marker is xDAT. This is true irrespective of MC transitivity.

(22)	Lardija	=ngga	dathina	maku
Y.	laɬic-a	=ŋka	ɬaɬin-a	maku
	wait_for-UNM		that-T	woman
	wait_for-UNM	3SG.NTRS.PRES[3SG.GOAL]	that[NTRS]	woman[NTRS]
	dathininja	dangkantha,	[warrajurluntha.]	
	ɬaɬin-ipca	ɬaŋka-ŋta	warac+ku[u-ŋta	
	that-XDAT	man-XDAT	go-XPROP-XDAT	
	that-GOAL	man-GOAL	go-FUT-COMP	
	‘That woman is waiting for the mani to Ø _i leave.’ [K5-1917A]			

Figure 5 shows these facts schematically, indicating the markers that appear on subjects and objects in both the MC and SC, as well as the SCCM marker on the subordinate clause. The co-reference conditions which require the use of SCCM are indicated by a double-headed arrow. (To avoid unnecessary complexity, the MC **CASE** markers shown are for full NP arguments only; as discussed in §4.4 TRMC subject pronouns differ).

Figure 5: SCCM and associated co-reference in complex sentences (Yukulta)

In complex sentences with TRMCS, the co-referential MC subject argument is inflected with XERGLOC and the same marker used for SCCM. Similarly, for NTRMCs the co-referential MC object argument is inflected with XDAT, which is the same marker used for SCCM, all of which leads SCCM to look rather like complementizing **CASE** agreement. However, notwithstanding the broad-brush appearance, a **CASE**-agreement analysis would not work entirely for the facts of Yukulta. SCCM employs XERGLOC and XDAT markers, even when the MC co-referent is a pronoun which is marked otherwise; and SCCM is blind to the distinction in NTRMCs between inverse and non-inverse objects. For example, in (23) co-reference is between the SC subject and an NTRMC inverse object. Although the inverse object is now marked with XERGLOC, the SCCM marker is still XDAT, just as in (22) where co-reference was with a regular, XDAT-marked NTRMC object.

(23)	Warlirra	=ngarrawa-nyi-nggi	ngarrawanji
Y.	waɭira	=ɲarawa-ni-ɲki	ɲa-ra-waɲ+ki
	NEG		1EXCL-DU-XOBL-XERGLOC
	NEG	1EXCL.DU.GOAL-2SG.NTRS-PST.IRR	1EXCL-DU-OBL-INV0
	gurri,	[ngaagawatharrbantha	wirdijarrbantha.]
	kuri	ɲa:kawat+ɲarpa-ɲta	wiɬic+ɲarpa-ɲta
	see	do_what-XPST-XDAT	stay-XPST-XDAT
	see[UMN]	do_what-PST-COMP	stay-PST-COMP
	‘You haven’t seen what we’ve been doing.’ (lit. ‘You haven’t seen us _i Ø _i what-doing’) [K3-1790B]		

To emphasize the point, Figure 6 shows the marking and co-reference relationships in complex sentences whose MC is nontransitive, for both regular and inverse NTRMCs.

Figure 6: SCCM and associated co-reference in complex sentences with nontransitive main clauses (Yukulta)

Given these facts, a reasonable analysis of SCCM in Yukulta is that it hinges not upon the inflectional **CASE** category of the MC co-referent — it is not simple complementizing **CASE** agreement — but on the co-referent’s grammatical role: co-reference with a TRMC argument in the TRS role triggers XERGLOC SCCM, while co-reference with a NTRMC argument in the GOAL or INV0 roles triggers XDAT SCCM. By the same token, given the broad similarity between SCCM markers and the **CASE** marking of MC co-referents, and given the Australian typological context, it is entirely conceivable that at an earlier diachronic stage, the SCCM system was a simple matter of complementizing **CASE** agreement, and hinged directly upon the main-clause co-referents’ **CASE**. In fact, let us now take a detour and consider what such a system would be like: a conceivable historical antecedent to Yukulta, with a straightforward complementizing **CASE** agreement system. The focus will be on how this would operate in complex clauses with nontransitive MCs.

5.3 SCCM in a plausible ancestral state

If SCCM operated as simple complementizing **CASE** agreement, at least in sentences with NTRMCs, then the diagram of marking and co-reference relationships for those sentences would appear as in Figure 7, that is, in sentences with NTRMCs, SCCM would be marked under some circumstances by XDAT, and under other circumstances by XERGLOC. The assumption will be that the triggers of the inverse are the same here as in Yukulta (§4.4). Let examine more carefully, then, what these ‘some’ and ‘other’ circumstance are, that govern the XDAT XERGLOC SCCM alternation.

Figure 7: Marking and co-reference in complex sentences with nontransitive main clauses (hypothesized ancestor of Yukulta)

In what follows I will step through a chain of reasoning, to bring to the fore an interesting and rather non-obvious curiosity of this system, namely that the contrast between XDAT and XERGLOC SCCM is largely, though not entirely, predictable from the **PERSON** values of the subject of the *subordinate* clause. It will be worth the reader's while to double check each step in the logical chain, to be satisfied that it is valid. The end-point in the chain is an association between SC subject **PERSON** and SCCM which turns out to be very similar to what we find in modern Kayardild.

To begin the chain of reasoning: in Figure 7 we see that the use of XDAT versus XERGLOC SCCM is triggered by the contrast that exists between regular and inverse NTRMCs. Accordingly, one could equally say that the factors that directly trigger the regular/inverse contrast now also indirectly trigger the SCCM contrast. The factors triggering the regular/inverse contrast are predominantly the **PERSON** values of the MC subject and MC object, with a minor role for object **NUMBER** (due to the second-on-first-singular idiosyncrasy). Therefore, we can also say that the SCCM contrast is ultimately triggered by MC subject **PERSON** and MC object **PERSON** and **NUMBER** values. Next however, consider that SCCM in these sentences is used when there is co-reference between the MC object and the SC subject, in which case the **PERSON** and **NUMBER** values of the MC object and the SC subject will be identical to one another. Consequently we could equally say — as odd as it might seem to do so — that the SCCM contrast is ultimately triggered by MC subject **PERSON** and SC *subject* **PERSON** and **NUMBER**.

This network of relationships can be set out in the manner of §2, as in Figure 8, where arrows show causal relationships and boxes contain sets of values that various parameters (or features) can take.

Figure 8: Causal links governing SCCM (hypothesized ancestor of Yukulta)

Figure 8 shows SCCM being ultimately determined by several factors. Working from the top, the ques-

		MC Object (= SC Subject)				
		1sg	1nonsg.excl	1incl	2	3
MC	1	<reflexive, or very rare in discourse>			XDAT	XDAT
Subject	2	XDAT	XERGLOC	XERGLOC	<reflexive>	XDAT
	3	XERGLOC	XERGLOC	XERGLOC	XERGLOC	XDAT

Table 4: Marking of objects in Yukulta nontransitive main clauses, according to subject and object properities; also, choice of SCCM marker according to the same factors in the hypothesized ancestor of Yukulta

SC subject:	1sg	1nonsg.excl	1incl	2	3
SCCM	XDAT ~ XERGLOC	XERGLOC	XERGLOC	XDAT ~ XERGLOC	XDAT

Table 5: Correlation between SC subject properties and SCCM (hypothetical ancestor of Yukulta)

tion of whether the MC object and SC subject are co-referential determines whether there is any SCCM marker at all; then, all other factors ‘flow through’ the mediating causal node of MC object **CASE**. Those other factors are MC subject **PERSON**, and the **PERSON** and **NUMBER** values of the co-referential MC object / SC subject. The next question I wish to examine is this: assuming that the MC object and SC subject are co-referential, how much can we predict about SCCM from SC subject properties alone? To begin answering this, Table 4 shows the combinations of MC subject **PERSON** and MC object / SC subject **PERSON** and **NUMBER** which trigger the use of XDAT and XERGLOC marking on NTRMC objects in Yukulta. In our hypothetical ancestral language with complementizing **CASE** agreement, these are also the triggers for the two SCCM markers.

Now, how much can be predicted purely from the SC subject? If the SC subject is **THIRD PERSON**, then the SCCM marker is entirely predictable; it is XDAT, irrespective of the MC subject (this is gleaned from Table 4 by looking vertically, down the rightmost column, corresponding to **THIRD PERSON** SC subject). Likewise, if the SC subject is **FIRST PERSON INCLUSIVE**, then the SCCM marker is entirely predictable; it is XERGLOC, irrespective of the MC subject. Continuing further, when the SC subject is **FIRST PERSON SINGULAR**, then the SCCM marker can be either XDAT or XERGLOC. When it is **FIRST PERSON EXCLUSIVE NON-SINGULAR**, then the SCCM is always XERGLOC, and when it is **SECOND PERSON**, then the SCCM marker can be either XDAT or XERGLOC. The correlations are summarized in Table 5.

Putting this together, Figure 8 highlighted the fact that this is a system where, if the SC subject and MC object are co-referential, then SCCM is ultimately predictable from MC subject **PERSON** and the MC object / SC subject **PERSON** and **NUMBER**. MC object **CASE** plays mediating role in the causal network. Secondly, we established just above that even if we ignore information about the MC subject, it is still possible to predict rather a lot about SCCM just from the properties of the SC subject. Keeping this in mind, in the next section we will examine the SCCM system of Kayardild. Kayardild lacks a MC inverse alternation. Consider then, what would happen if we took the hypothetical system we have been discussing in this section and removed the MC inverse system. The initial effect on the causal network would be to remove the mediating node at its centre, leaving Figure 9.

Figure 9 represents something like what history would hand to learners of a language which previously had been as in Figure 8, but had just lost its main clause inverse system. We can note several things. First, the use versus non-use of SCCM is still transparently related to co-reference between

Figure 9: Causal network of Figure 8, but with the mediating node removed.

SC subject and MC object: it is used only when they are co-referential. This would be a correlation that learners would apprehend with ease. Second, we established above that a large amount of the variation between the use of XDAT and XERGLOC as the SCCM marker is correlated with properties of the SC subject. It is highly likely that learners would also apprehend these correlations, between SC subject properties and SCCM marking, since they are all properties of the SC. The next question is, would learners also detect the complex causal link between MC subject properties and SCCM? Perhaps so, but even if that link were acquired, it would come under strong pressure to be simplified, and either be eliminated from the system, or replaced by something else, by some third factor. Supposing it were replaced by a third factor, what remains is Figure 10.

Figure 10: Causal network of Figure 9, relinked and simplified.

5.4 SCCM in Kayardild

In Kayardild, there are two SCCM markers, XDAT and XERGLOC (cf. examples in §1). The use versus non-use of SCCM is related to co-reference between arguments in the SC and MC. The choice between XDAT and XERGLOC as the SCCM marker is primarily determined by SC subject **PERSON**, though there is a third factor, namely **SOLIDARITY**: whether ‘the speaker wants to group him/herself with the addressee’. That is, Kayardild conforms very closely to what we predicted, just above, would result if a language with a Yukulta-like MC inverse system and complementizing **CASE** agreement were to lose its MC inverse system. Kayardild, as mentioned earlier, lacks a MC inverse system.

There are some differences between Kayardild and the exact system we considered at the end of §5.3. First, Kayardild’s SCCM is triggered by co-reference not only between SC subject and MC object, but between any combination of SC and MC subject and object other than subject–subject (Evans

1995a:489–90).⁹ For example, sentence (24)¹⁰ shows the use of SCCM due to co-reference between SC and MC objects.

(24)	Nyingka	kurrijarr	ngijinjina	banganaa,
K.	niŋ-ka	kuric-ŋara	ŋicijŋ+kina:	paŋa+kina:
	2SG-T	see-XPST	1SG.XOBL-XABL	turtle-XABL
	2SG.SBJ	see-PST	1SG.POSS-PST[TRO]	turtle-PST[TRO]
	[ngijuwa warijarrmatharranth?]			
	ŋicu-wa wa.ŋicarmat-ŋara-ŋta			
	1SG.XOBL-T	bring_ashore-XPST-XDAT		
	1SG.COMP1[SBJ]	bring_ashore-PST-COMP1		
	'Did you see my turtle _i , that I caught Ø _i ?' [W1960]			

Second, the match between SC subject properties and the use of XDAT versus XERGLOC is not precisely the same as in §5.3, although most idiosyncrasies do match exactly: the only differences are in the FIRST EXCLUSIVE. In Kayardild, when the SC subject is FIRST **PERSON** SINGULAR, then the SCCM marker is XDAT only. In Table 4, it was XDAT or XERGLOC. And in Kayardild, when the SC subject is FIRST **PERSON** EXCLUSIVE NON-SINGULAR, then the SCCM is always XDAT; in Table 4, it is XERGLOC.

My conclusion from this is as follows. It is very likely that Yukulta and Kayardild share a common ancestor which had a MC inverse system and complementizing **CASE** agreement. On the Yukulta branch of the family tree, complementizing **CASE** agreement was restructured, resulting (in complex sentences with a NTRMC) in the loss of the earlier alternation between XDAT and XERGLOC SCCM and retention of the regular (non-inverse) XDAT alternant, which would have been more common in discourse. The inherited MC inverse system was retained though. On the Kayardild branch of the tree, the MC inverse system was lost, and the SCCM causal network restructured. The restructuring involved the co-opting of **SOLIDARITY** to predict a residual, inherited alternation between XDAT and XERGLOC SCCM in SCs with SECOND **PERSON** subjects. In addition, the SCCM marker for all FIRST EXCLUSIVE SC subjects was regularized to just one form, XDAT. This may have occurred through a generalization of the XDAT marker, which was already associated with FIRST SINGULAR, the most frequent FIRST EXCLUSIVE category in discourse. The causal network for modern Kayardild is shown in Figure 11.

Figure 11: Causal network of SCCM in Kayardild

There is one outstanding question. In Yukulta, there is one kind of SCCM marking used in sentences

⁹One exception appears to be sentences whose matrix clause is a 1INCL imperative, in which case SCCM may be used even when the only co-reference is between SC and MC subject; see Round (2017) regarding imperatives in Tangkic, and Evans (1995a:508–514) for other idiosyncrasies in Kayardild’s triggers of SCCM.

¹⁰In (24) I use ‘COMP1’ in the gloss for comparability with examples (1)–(4). A more informative gloss, consistent with the analysis developed above, would be ‘1EXCL.SUBJ’.

with TRMCs, and another kind used in sentences with NTRMCs. Since §5.3, the discussion has only been about the latter, in terms of which we can account neatly for the modern Kayardild system. But why does Kayardild SCCM relate only to Yukulta’s NTRMCs system, and not its TRMC system? Is there any evidence, other than in its SCCM *xDAT* ~ *XERGLOC* alternation, that pre-Kayardild lost the system of SCCM associated with TRMCs? The answer is yes, and there is both language-internal and comparative evidence.

The comparative evidence will come from Lardil, which we consider next. The language-internal evidence relates to co-reference. In Yukulta sentences with TRMCs, SCCM is triggered by co-reference between the subjects of the SC and MC. If pre-Kayardild retained some vestige of this system, we might expect it to continue this criterion of subject–subject co-reference. However, as mentioned earlier in this section, that is the one condition under which SCCM is *not* triggered in Kayardild. Thus in Kayardild, the idiosyncrasies of the SCCM *xDAT* ~ *XERGLOC* alternation, and the criteria on co-reference point in the same direction, suggesting that pre-Kayardild entirely lost the equivalent of Yukulta’s SCCM system associated with TRMCs.

In this section I have argued that Kayardild is one short historical step away from an ancestral language which possessed a MC inverse system and complementizing **CASE** agreement — a language which is also one short historical step away from Yukulta, and which I reconstruct as ancestral to both.

5.5 SCCM in Lardil

In Lardil, there is just one SCCM marker, *xDAT*. Its use is triggered by co-reference between SC subject and MC object as in (25) or, in subordinate topicalized-object constructions (Klokeid 1976; Round 2017), between SC topicalized-object and MC object, as in (26).

(25)	Ngada	rathur	kiinkur	ngawur,
L.	ɲata	ɭac+kuɭu	ki:n+kuɭu	ɲawu+kuɭu
	1SG	spear-XPROP	that-XPROP	dog-XPROP
	1SG.SBJ	spear-FUT	that-FUT[TRO]	dog-FUT[TRO]

[ngithunarrba	betharrbanu.]
ɲitun+ɲarpa-iɲta	peɬ+ɲarpa-iɲta=u
1SG-XPST-XDAT	bite-XPST-XDAT=ACT
1SG-PST-COMP[TRO]	bite-PST-COMP=ACT
‘I’m going to spear the dog _i , which Ø _i bit me.’ [Klokeid 1976:461]	

Co-reference	Yukulta (TrMC)	Yukulta (NtrMC)	Lardil	Kayardild
SC Subj & MC Subj	+			
SC Subj & MC Obj		+	+	+
SC Obj & MC Obj			+	+
SC Obj & MC Subj				+

Table 6: Co-reference triggers of SCCM

(26)	Ngada	warnawuthur	kiinkur	wurdaljiwur
L.	ɲata	waɲawuc+ku.ɲu	ki:n+ku.ɲu	wuɬalci+ku.ɲu
	1SG	cook-XPROP	that-XPROP	meat-XPROP
	1SG.SBJ	cook-FUT	that-FUT[TRO]	meat-FUT[TRO]
	[ngimbenin	wirimatharrba.]		
	ɲiɲ+pæn-iɲɬa	wi.ɲimatɬ+ɲarpa-iɲɬa		
	2SG-XOBL-XDAT	bring-XPST-XDAT		
	2SG-SBJ-COMP	bring-PST-COMP		
	'I'll cook that food _i , that you brought [Ø _i] _{Topic} .' [Klokeid 1976:385]			

In this sense, the co-reference triggers in Lardil sit part way between those of Kayardild and those of Yukulta's sentences with NTRMCs, as summarized in Table 6.

Like Kayardild, Lardil's co-reference triggers of SCCM correspond only to those of Yukulta's NTRMC sentences, not its TRMC sentences. Likewise, its SCCM marker xDAT corresponds only to SCCM in Yukulta's NTRMC sentences, not its TrMC sentences. And finally, it will be recalled from §4.3, that the inflectional properties Lardil's PRESENT **TENSE** MCs also correspond to Yukulta's NTRMCs but not its TRMCs. Thus, while Lardil and Kayardild retain multiple points of morphosyntactic cognacy the Yukulta's NTRMC, they both appear to have undergone the wholesale loss of morphosyntax that is cognate with Yukulta's TRMCs.

5.6 Whence Kayardild's syntax?

The lexicon of Kayardild is very similar to Yukulta, rather than Lardil. Even its inflectional forms and their allomorphy are more like those of Yukulta. Yet the functions that are served by its inflectional forms, and the syntax of its sentences are closer to Lardil. Explaining precisely how this came to be is a complex task, and beyond the remit of this chapter. However, I wish to make one observation that stems from the discussion above.¹¹

Suppose, rather simplistically, that a language with a MC inverse system and complementizing **CASE** agreement, as discussed in §5.3, co-opted the main clause syntax of Lardil, while largely keeping its own lexicon and morphological forms. The result would be the loss of the old TRMC system and loss of the MC inverse system. The alternating xDAT XERGLOC SCCM with a causal network as in Figure 8 would lose its mediating node. Restructuring could ensue, and the result could quite plausibly be something like Kayardild. In the MC, objects would become **TENSE**-marked, largely mirroring **TENSE** marking in SCs; perhaps there would even be over-generalization, resulting in **TENSE**-marked

¹¹Many thanks to Nick Evans for discussion of these thorny questions.

PRESENT **TENSE** MC objects (found in Kayardild but not Lardil). Clearly, this is far from a full reconstruction, but it illustrates, I hope, that if recent ancestor of Lardil came into heavy contact with a language like that of §5.3, then an outcome like Kayardild has *prima facie* plausibility. By the same token however, another route could deliver us to the same destination.

Suppose instead, that Lardil itself traces back to a language as in §5.3, with a MC inverse system and complementizing **CASE** agreement. Suppose then that its own internal changes led to the generalization of old NTRMC morphosyntax at the expense of TRMCs, leading to the loss of the latter; and that it lost the old MC inverse system. This too, could remove the mediating node from a causal network like Figure 8 and lead to the emergence, in pre-Lardil, of alternating XDAT~XERGLOC SCCM. Now, if this system came into heavy contact with a Yukulta-like language and was extensively re-lexified, the result would resemble Kayardild. Meanwhile, if the same system, in another dialect, lost its XDAT~XERGLOC SCCM alternation by generalizing the XDAT marker, it would resemble Lardil.

The point here seems to be firstly, that some kind of intense language contact event is likely to have been involved, but secondly that it may have involved the borrowing of Lardil-like syntax into a language like in §5.3, or the borrowing of Yukulta-like lexicon into a language whose syntax resembled Lardil but with an XDAT~XERGLOC SCCM alternation. Finding evidence which distinguishes between these two scenarios is likely to be challenging, but there may be a lead here worth following up in future research. Namely, (i) if we assume that proto Tangkic had a basic contrast between TRMCs and NTRMCs largely like in Yukulta, then logically, Lardil must have lost the old TRMC syntax at some point in its past; however, (ii) if proto Tangkic had a MC inverse system, then Lardil must have lost that too, and as I argued just above in my second scenario, loss of TRMCs and the MC inverse system is enough to trigger the creation of a Kayardild-like XDAT~XERGLOC SCCM alternation as an internal change in pre-Lardil, historically prior to the contact event that gave rise to the lexical-syntactic mix of pre-Kayardild. On the other hand, if proto Tangkic did not have a MC inverse system, and it was innovated in a shared ancestor of Yukulta and Kayardild, then the XDAT XERGLOC SCCM alternation cannot have arisen internally in pre-Lardil since its preconditions would not have existed there; it must have arisen in pre-Kayardild, and quite possibly, then, as a consequence of the contact event. Interestingly, the fact that modern Lardil lacks an XDAT XERGLOC SCCM alternation is not diagnostic, since it could easily have lost it (and note that in both scenarios above, such an alternation must have been lost in pre-Yukulta). Consequently, a key question for Tangkic historical research is, what evidence is there that the MC inverse system was present in proto Tangkic? This is work for the future.

6 Shifts in the systems of agreement

We turn now from the mechanics of Tangkic reconstruction, to its theoretical implications. Irrespective of whether it occurred in pre-Lardil or solely in pre-Kayardild, I have argued that the XDAT~XERGLOC SCCM alternation found today only in Kayardild resulted from the re-structuring of a causal network whose mediating node was removed by the loss of a MC inverse system of a kind attested directly in modern Yukulta. One assumption made, was that in the antecedent system there was complementizing **CASE** agreement, so that the XDAT XERGLOC marking on SCs was a matter of agreement in **CASE** between the subordinate clause, which in Corbett's (2006) terms was the agreement target, and a co-referential MC object, which was the agreement controller. The notion that the antecedent system could have possessed this kind of complementizing **CASE**

agreement draws its plausibility from three sources. First, although modern Yukulta does not have complementizing **CASE** agreement *per se*, its superficial inflectional markings are very similar to one. Second, only by assuming that the ancestral system was like this can we explain in considerable detail the empirical idiosyncrasies of Kayardild's XDAT XERGLOC marking and their correlation with **PERSON** values. Third, complementizing **CASE** agreement of this kind is typologically unremarkable in the Australian context.

In the antecedent system therefore, the XDAT~XERGLOC marking on SCs was an exponent of **CASE**, but the modern Kayardild system is quite different. The choice of SCCM marker in Kayardild is determined primarily by subject **PERSON** with slight conditioning by **SOLIDARITY**, and the choice of whether any SCCM is used is determined by co-reference. In Corbett's (2006) terms, this too is agreement. The subordinate subject is the controller, all of the words of the subordinate clause are targets, the feature agreed for is **PERSON**, and there are two conditions on that agreement, the factors of co-reference and **SOLIDARITY**. Because of its historical source though, this **PERSON**-agreement system is highly non-canonical. Its marker is etymologically **CASE**, the location where it is marked is the entire subordinate clause, and it contrasts only two surface categories despite having four underlying values. All of these non-canonical traits are due to its immediate history, as a system of complementizing **CASE** marking in subordinate clauses, in agreement with main-clause objects that took two different **CASE** values.

7 Typology with and without diachrony

The reconstruction in §5 is a novel contribution to Tangkic diachronic morphosyntax. Previous synchronic analyses of Kayardild SCCM did not have the benefit of it, and consequently they provide a retrospective case study in how challenging the typology of *rara* can be in the absence of a diachronic backstory.

Evans (1995a) analyses modern Kayardild SCCM as complementizing **CASE** marking. As a starting point this is not unreasonable. Due to its recent history, SCCM in Kayardild has many traits in common with true complementizing **CASE** systems. However, as Evans (2003) observes, it is unclear how one should account theoretically for such complementizing **CASE** inflection, since if it is a matter of **CASE** agreement, then the agreement controller appears to be missing; and if it is a matter of **CASE** government, then the governor appears to be missing. Given the diachronic context we can see why this is so. The causal network of true complementizing **CASE** in the antecedent language system (Figure 8) was firstly impoverished when the MC inverse system was lost (Figure 9), removing the erstwhile controller of agreement, and then significantly rewired in the change to Kayardild (Figures 10, 11). The modern system is no longer a system of complementizing **CASE**, notwithstanding its superficial resemblances. In the absence of that diachronic context, Evans (2003) concludes that Kayardild poses problems for theories of agreement.

Round (2013) approaches Kayardild SCCM from another angle. Whereas Evans' analysis is in terms of a typologically not-uncommon feature (complementizing **CASE**), and leads to problems accounting for the inflectional mechanism, Round (2013) provides a straightforward inflectional mechanism (feature percolation) but arrives at an analysis with two typologically unusual features. The first, **COMPLEMENTIZATION**, corresponds in causal terms to the node 'COREFERENCE' in Figure 11: it is a feature that is introduced into the clause under the same conditions that trigger SCCM. The second feature is **SEJUNCT**, a purposefully unusual name for what Round (2013) argues is an unusual feature:

one associated directly with first person exclusive, second person (non-solidarity), and third person. In causal terms, **SEJUNCT** corresponds to the merger of the two nodes ‘**SOLIDARITY**’ and ‘SC Subj **PERSON**’ in Figure 11. Round (2013) compares Kayardild SCCM to the typology of **PERSON** marking, but having observed that SCCM has just two surface oppositions, is marked across a whole clause, and intersects with the pragmatics of **SOLIDARITY**, the conclusion is that **SEJUNCT** is so unlike other synchronic **PERSON** systems, that to call it **PERSON** would be to mischaracterize it. Again, the diachronic context clarifies how the analyst reached this point. **PERSON** agreement in Kayardild is indeed very different to other **PERSON** agreement systems, because it has just evolved by a highly infrequent route, from a complementizing **CASE** system fed by an inverse system.

Neither Evans (1995a; 2003) nor Round (2013) can be faulted with poor argumentation or poor typologizing. Evans compared Kayardild SCCM to typological systems that resembled it in terms of morphological forms and their distributions (complementizing **CASE**), and then noted that the mechanism was odd. Round compared SCCM to typological systems that resembled its function (**PERSON** agreement) and got the mechanism to work, but noted that the morphological forms were odd (in terms of syncretism and location). Both clearly signaled what was unusual. And this is the point at which observations about *rara* often stop (Cysouw and Wohlgemuth 2010). However, with the diachronic analysis in hand these issues can be resolved. Not only can we now explain the particulars of Kayardild SCCM and its apparent typological mismatches, but we can also characterize the system as **PERSON** agreement with conditioning by solidarity and co-reference, in the knowledge that there is good reason why a **PERSON** system might look like this. In the absence of the diachronic analysis, the act of declaring Kayardild SCCM to be **PERSON** marking may have resembled hammering a square peg into a round hole, but with it we have both resolved a quandary that Kayardild had posed, and extended our understanding of the typology of **PERSON** systems.

8 The scientific response to *rara*: examine diachrony

Evans and Round both observed that the marking of complementized clauses in Kayardild constitutes a *rarum*, and identified where it clashed with our current understanding of empirical synchronic typological facts and our theories for explaining them. At face value, the *rarum* presented a challenge to theories such as Corbett’s theories of agreement (Corbett 2006) and features (Corbett 2012), which propose that the variety of inflectional mechanisms and features available to human languages is circumscribed: Evans’ analysis challenged the former, and Round’s the latter. There are two broad responses to a state of affairs such as this, and the debate between them is ongoing. One is to doubt whether it is possible to circumscribe the variety of human language (Evans and Levinson 2009). The other is to persist with the task of composing empirical insights in into circumscribed theories of language, embracing apparent counter-examples as a prompt both to revise those theories, and to seek avenues that bring improved insights into the empirical facts. Here I have sought to demonstrate that diachrony is one such avenue.

There are well-worn arguments that we should not underestimate the utility of diachrony for synchronic typologizing (Greenberg 1966; Greenberg 1978; Bybee 1988; Blevins 2004, inter alia). To take a whimsical example, if the world contained 100,000 languages, I would confidently predict that we would encounter a rare but recurring pattern of (i) subject **PERSON** agreement, (ii) found in subordinate but not main clauses, (iii) marked by etymological **CASE** markers with (iv) only two formal

distinctions, in which (v) the most common pattern is that third person differs from first person, and second person is syncretic with either first or third person or both, and is conditioned (vi) by co-reference, and perhaps (vii) by some other factor in the second person. That is a terribly specific synchronic typological claim to make, and the fact I have any confidence in it is by virtue of diachrony. Of course, there are not 100,000 languages and so this generalization is highly unlikely to emerge from synchronic observations alone. And that is the problem of *rara*: given our necessarily limited sample from the vast number of possible human languages, there very likely are many correct, synchronic typological generalizations which typologists will be capable of discovering only through diachrony.

Appendix. Kayardild Topic and Focus constructions

Kayardild has two complementized clause constructions in which the marker of complementization does not appear on every word of the clause (Round 2013). Example (27) illustrates a Topic construction. The clause is complementized and words carry **PERSON** agreement with the first person exclusive subject (glossed 1SBJ), however the Topic DP escapes **PERSON** agreement.

- (27) Ngijuwa kuluujuuntha birr.
 K. ŋicu-wa kulu:c+ku:-ŋta pi-r-a
 1SG.XOBL-T scratch-XPROP-XDAT 3-PL-T
 1SG.1SBJ[SBJ] scratch-FUT-1SBJ 3-PL[TOPIC.TRO]
 ‘I’ll scratch them.’ [W1960]

In Kayardild, all inflectional features have a syntactic domain within which they appear. The Topic construction can be accounted for by assuming that the Topic DP is outside the domain of SCCM **PERSON** agreement (Round 2013). More complicated is the Focus construction, as in (28). Here, the clause is complementized and words carry **PERSON** agreement with the third person subject (glossed 3SBJ). The Focus DP however lacks this XDAT inflection and is marked with XERGLOC.

- (28) Dangkawalathiya dalinkiya
 K. taŋka+palat+ki talic-n+ki
 person-PL-XERGLOC come-XNOMZ-XERGLOC
 person-PL-LOC[FOCUS.SBJ] some-NOMZ-LOC[FOCUS.SBJ]
- warrajuuntha wurankiiwathuunth.
 wara-c+ku:-ŋta wu.ɬan+ki:waɬ+ku:-ŋta
 go-XPROP-XDAT food-COLL-XPROP-XDAT
 go-FUT-3SBJ food-COLL-FUT-3SBJ
 ‘The people coming are going for food’ [R2005-jul21]

In Kayardild, there are several feature values realised by XERGLOC, and we can ask which one it is that is being realised on Focus DPs. Round (2013:95–96) presents diagnostic evidence suggesting that the XERGLOC on Focus DPs is neither **TENSE** nor **CASE**. The argument regarding **TENSE** is secure, but the argument regarding **CASE** hinges on one example, for which another analysis is possible.¹² I propose

¹²In a two word example Ngada rarungki! ‘I’m in the south!’, Round (2013:96) observes that the pronoun ngada ‘I’ has no overt inflection while the word rarungki ‘in the south’ has XERGLOC inflection. Both are taken to be Focus DPs. Now, in complementized clauses with XERGLOC SCCM, pronouns are uninflected (like ngada is), whereas pronouns inflected with LOCATIVE **CASE** have overt inflection (unlike ngada). So, Round concludes, the lack of inflection on the Focus DP

here, then, that Focus DPs are inflected with **LOCATIVE CASE**, bringing my analysis back closer to Evans' (1995a:141–142) analysis of bare Focus DPs, that occupy a clause all on their own. It also simplifies the analysis of Focus constructions relative to the analysis in Round (2013), which used two features associated with complementization. The analysis here is that there is only the one feature associated with complementization, namely **PERSON** agreement. Focus DPs, like Topic DPs, are outside of the syntactic domain associated with that feature, and do not inflect for it. However, in contrast to Topic DPs which are not inflected at all, Focus DPs are inflected for **LOCATIVE CASE**.

References

- Anderson, Stephen R. 1981. Why phonology isn't "natural". *Linguistic inquiry* 12(4). 493–539.
- Anderson, Stephen R. 1988. Morphological change. In David Sankoff & Frederick J. Newmeyer (eds.), *Linguistics: the cambridge survey*, 324–362. Cambridge: Cambridge University Press.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Aronoff, M. 1994. *Morphology by itself: stems and inflectional classes*. Cambridge, MA: The MIT Press.
- Bach, Emmon & R. T. Harms. 1972. How do languages get crazy rules? In R.P. Stockwell & R. K. S. Macaulay (eds.), *Linguistic change and generative theory*, 1–21. Bloomington: Indiana University Press.
- Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Brown, Dunstan, Marina Chumakina & Greville G. Corbett. 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Bybee, Joan L. 1988. The diachronic dimension in explanation. In John A Hawkins (ed.), *Explaining language universals*, 350–379. Oxford: Basil Blackwell.
- Cook, Jacqueline. 2017. *A sketch grammar of yangkaal* dissertation.
- Cook, Jacqueline & Erich R. Round. 2017. Draft yangkaal dictionary. University of Queensland, Brisbane.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S Rood (eds.), *Linguistic diversity and language theories*, 25–49. Amsterdam: John Benjamins.
- Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2007. Deponency, syncretism, and what lies between. In Matthew Baerman et al. (eds.), *Deponency and morphological mismatches*, 21–44. Oxford: Oxford University Press.
- Cysouw, Michael & Jan Wohlgemuth. 2010. The other end of universals: theory and typology of rara. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: how rarities affect linguistic theory*, 1–10.

ngada can be explained in terms of the usual SCCM feature and not **CASE**, and this suggests that the XERGLOC marker on all Focus DPs is a realization of the SCCM feature, not **CASE**. However, this misses another possible analysis of the key example. Directional stem like rarung- do not inflect for **LOCATIVE CASE** or for the 'ACTUAL' present **TENSE** (both marked by XERGLOC), at least, they do not do so if they are the head of the NP. But they can inflect if they are modifiers in a NP, and it is possible that rarunki here is a modifier of a headless NP, which are not uncommon in Kayardild (Round 2013:158–167). In that case, neither ngada nor rarunki need to be analysed as Focus DPs. Ngada is a normal subject DP and rarunki is a modifier in a headless predicate DP which is inflected either for **CASE** or **TENSE**. The literal meaning of the utterance would be 'I'm in the south(ern land)', i.e., on Bentinck Island.

- Daniel, Michael. 2011. Linguistic typology and the study of language. In Jae Jung Song (ed.), *The oxford handbook of linguistic typology*, 43–68. Oxford: Oxford University Press.
- Dench, Alan & Nicholas Evans. 1988. Multiple case-marking in australian languages. *Australian Journal of Linguistics* 8(1). 1–47.
- Dixon, R. M. W. 1979. Ergativity. *Language* 55(1). 59–138.
- Evans, Nicholas. 1992. *Kayardild dictionary and thesaurus*. Department of Linguistics & Language Studies, University of Melbourne.
- Evans, Nicholas. 1995a. *A grammar of kayardild, with historical-comparative notes on tangkic*. Berlin: Walter de Gruyter.
- Evans, Nicholas. 1995b. Multiple case in kayardild: anti-iconic suffix ordering and the diachronic filter. In Frans Plank (ed.), *Double case: agreement by suffixaufnahme*, 396–428. Oxford: Oxford University Press.
- Evans, Nicholas. 2003. Typologies of agreement: some problems from kayardild. *Transactions of the Philological Society* 101(2). 203–234.
- Evans, Nicholas. 2005. East across a narrow sea: micro-colonization and synthetic prehistory in the wellesley islands, northern australia. In Toshiki Osada (ed.), *Linguistics, archaeology and the human past*. 9–39. Kyoto: Research Institute for Humanity & Nature.
- Evans, Nicholas & Steven Levinson. 2009. The myth of language universals : language diversity and its importance for cognitive science. *Behavioral & Brain Sciences* 32(5):429-448, 472-492.
- Evans, Nicholas & Erich R. Round. 2017. *Kaiadilt: kayardild dictionary app*. Canberra: Centre of Excellence for the Dynamics of Language.
- Gildea, Spike & Fernando Zúñiga. 2016. Referential hierarchies: a new look at some historical and typological patterns. *Linguistics* 54(3). 483–529.
- Givón, Talmy. 1979. *On understanding grammar*. Orlando: Academic Press.
- Goddard, Cliff. 1982. Case systems and case marking in australian languages: a new interpretation. *Australian Journal of Linguistics* 2(2). 167–196.
- Greenberg, J. H. 1978. Diachrony, synchrony, and language universals. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, volume 1, method and theory*. 61–91. Stanford, California: Stanford University Press.
- Greenberg, Joseph H. 1966. Synchronic and diachronic universals in phonology. *Language* 42(2). 508–517.
- Grossman, Eitan. 2016. From rarum to rarissimum: an unexpected zero person marker. *Linguistic Typology* 20(1). 1–23.
- Hale, Kenneth L. 1973. Deep-surface canonical disparities in relation to analysis and change: an australian example. In T. A. Sebeok (ed.), *Current trends in linguistics 8: linguistics in oceania*, 401–458. The Hague: Mouton de Gruyter.
- Harris, Alice C. 2008. On the explanation of typologically unusual structures. In Jeff Good (ed.), *Linguistic universals and language change*, 54–76. Oxford: Oxford University Press.
- Harris, Alice C. 2010. Explaining typologically unusual structures: the role of probability. In Michael Cysouw & Jan Wohlgemuth (eds.), *Rethinking universals: how rarities affect linguistic theory*. 91–103.
- Hartwig, Luke. 2014. *Case marking and clause function in tangkic languages: a comparative morphological study* dissertation.
- Hopper, Paul J. & S. A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.

- Jacques, Guillaume & Anton Antonov. 2014. Direct/inverse systems. *Language and linguistics Compass* 8(7). 301–318.
- Keen, Sandra. 1970. Materials in the yukulta language, 1968-1970.
- Keen, Sandra L. 1983. Yukulta. In R. M. W. Dixon & B. J. Blake (eds.), *Handbook of australian languages, volume 3*, 190–304. Canberra: ANU Press.
- Klokeid, Terry J. 1976. *Topics in lardil grammar* dissertation.
- Leman, Ngakulmungan Kangka. 1996. *Lardil dictionary: a vocabulary of the language of the lardil people, mornington island, gulf of carpentaria, queensland; with english-lardil finder list*. Mornington Shire Council.
- Memmott, Paul et al. 2016. Fission, fusion and syncretism: linguistic and environmental changes amongst the tangkic people of the southern gulf of carpentaria, northern australia. *Land and Language in Cape York Peninsula and the Gulf Country. Culture and Language Use*. 105–136.
- Nancarrow, Cassy (ed.). 2014. *Ganggalida dictionary and thesaurus: a vocabulary of the yugulda language of the ganggalida people, north-west queensland*. Cairns: Carpentaria Land Council Aboriginal Corporation.
- Nordlinger, Rachel. 2014. Constituency and grammatical relations in australian languages. In Harold J. Koch & Rachel Nordlinger (eds.), *The languages and linguistics of australia*, 215–262.
- Nordlinger, Rachel & Louisa Sadler. 2004. Nominal tense in crosslinguistic perspective. *Language* 80(4). 776–806.
- Plank, Frans. 2000. *Das grammatische Raritätenkabinett: A leisurely collection to entertain and instruct*. Konstanz: Universität Konstanz. <http://typo.uni-konstanz.de/rara/intro/index.php>.
- Richards, Norvin. 2013. Lardil “case stacking” and the timing of case assignment. *Syntax* 16(1). 42–76.
- Round, Erich R. 2013. *Kayardild morphology and syntax*. English. Oxford: Oxford University Press.
- Round, Erich R. 2005. *A first Kayardild audiovisual text corpus, with prosodic annotations: End of award report (FTG0025)*. Tech. rep. New Haven, CT.
- Round, Erich R. 2007. *Linguistic and ethnographic documentation of Kayardild: End of award report (IGS0039)*. Tech. rep. New Haven, CT.
- Round, Erich R. 2009. *Kayardild phonology, morphology, and morphosyntax*. New Haven dissertation.
- Round, Erich R. 2011. Morphemes as a level of representation capture unity of exponence across the inflection-derivation divide. *Linguistica* 51. 217–230.
- Round, Erich R. 2014. A description of ganggalida grammar. In Cassy Nancarrow (ed.), *Ganggalida dictionary and thesaurus: a vocabulary of the yugulda language of the ganggalida people, north-west queensland*, 173–205. Cairns: Carpentaria Land Council Aboriginal Corporation.
- Round, Erich R. 2015. Rhizomorphemes, meromorphemes, and metamorphemes. In Greville G. Corbett, Dunstan Brown & Matthew Baerman (eds.), *Understanding and measuring morphological complexity*. 29–52. Oxford: Oxford University Press.
- Round, Erich R. 2016. Kayardild inflectional morphotactics is morphomic. In Ana Luis & Ricardo Bermúdez-Otero (eds.), *The morpheme debate*, 228–247. [Online; accessed 2017-01-30]. Oxford: Oxford University Press.
- Round, Erich R. 2017. The tangkic languages of australia: phonology and morphosyntax of lardil, kayardild and yukulta. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. [Online; accessed 2017-05-03]. <http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-159>.

- Round, Erich R. & Greville G. Corbett. 2017. The theory of feature systems: one feature versus two for kayardild tense-aspect-mood. *Morphology* 27(1). 21–75.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical categories in australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Wang, William S.Y. 1968. Vowel features, paired variables, and the english vowel shift. *Language*. 695–708.
- Witzlack-Makarevich, Alena et al. 2016. Decomposing hierarchical alignment: co-arguments as conditions on alignment and the limits of referential hierarchies as explanations in verb agreement. *Linguistics* 54(3). 531–561.
- Wurm, Stephen A. 1960. Recordings made in the gayadilt language of bentinck island.