Diversity in the Numeral Systems of Australian Languages

CLAIRE BOWERN

Yale University

JASON ZENTZ

Yale University

Abstract. While the numeral systems of Australian languages are small, they are not uniform. In this article we shed light on the extent of variation in small numeral systems by systematically surveying 189 languages from Pama-Nyungan and non-Pama-Nyungan families. We show that, contra previous assumptions, Australian languages vary extensively in the limits of their numeral system, in the ways in which smaller numerals may be combined to form larger ones, and in their ability to denote vague quantities like 'some' or 'few'. We also investigate the etymological sources of words for numerals and find that few terms are reconstructible.

1. Introduction. Australian languages tend to have restricted numeral systems (Dixon 1980:107–8), stereotypically "one, two, three, many." Despite the prevalence of small numeral systems, not only in Australia, but in other parts of the world, they remain understudied. They are often excluded from discussion in surveys of larger systems on the grounds that they have different properties, even though these properties have yet to be systematically investigated.[1] Hanke (2010:64), for example, excludes Australian languages from his typology because the systems are no longer used and are poorly described (though he provides no references for these statements). Australian numeral systems are also widely assumed to be universally simple. For example, Hanke says that "the omission of Australian languages [from his study] is explained by the fact that with the exception of borrowings, almost all have systems of minimal complexity: after 1 and 2 follow quantifiers for 'a few' and 'many'; 3 seldom occurs (cf. Lynch 1998:249)" (2005:17).[2]

Hanke (like Lynch earlier) is here closely repeating a statement from Dixon (1980:107–8), who states that Australian languages lack numeral systems entirely, exhibiting instead a set of determiners marked for grammatical number. Dixon credits Kenneth Hale[3] with the idea that numeral-like items in Australian languages may be indefinite determiners. However, such a view belies the variation in even small systems across the continent, which goes far beyond that seen crosslinguistically in quantifier or determiner systems.

Here, we examine properties of Australian numeral systems and etymology of forms to shed light on the processes involved in the creation and maintenance of these small systems. Findings here are based on a sample of 189 Australian

133

languages, representing thirteen families and three isolates. We find counter-examples to several claims in the literature (Greenberg 1978:248–92; Hammarström 2008) and discuss data that have too often been excluded from previous studies (e.g., Hanke 2010:64). We find that many of the generalizations made about Australian systems—beyond the smallness of the system—are false; for example, terms for 'three' are far from rare, occurring in 94 percent of the languages sampled. We also find considerable variation in the semantics of the numeral terms. Given the current interest in the implication of small numeral systems for cultural and cognitive constraints on language (Evans and Levinson 2009), a review of such systems and their properties is timely. We discuss the variation in Australian numeral systems in terms of the range of such systems, how numerals are combined, whether the numerals may denote vague or inexact quantities, and etymological sources of numeral words.

2. Overview of the data.   The data for this survey were drawn from Bowern's comparative lexical files of Pama-Nyungan languages,[4] along with additional data from non-Pama-Nyungan languages. While we rely primarily on lexical data from dictionaries and wordlists for this survey, we have supplemented that data where possible by information from reference and sketch grammars. The database was searched for all varieties where translations for numerals were recorded. We selected languages for which we could be fairly sure that the number of numerals was likely to represent the extent of the system, and not be the result of gaps in recording. Where multiple dialects of the same language were in the database, they were included as separate varieties; for example, Gamilaraay, Yuwaalaraay, and Yuwaaliyaay (Ash, Giacon, and Lissarrague 2003) are included separately. Numeral systems show some (minor) dialect variation, so all available varieties were included so as not to privilege one system over another. The survey included 189 varieties; these are called "languages" in the rest of the article. Map 1 gives the locations of the languages in the survey.

Most of the surveyed languages are from the Pama-Nyungan family. Twenty-one Pama-Nyungan subgroups are represented; this includes almost all the major subgroups identified in the family (see, e.g., O'Grady, Voegelin, and Voegelin 1966).[5] Also included are the isolates Tiwi, Limilngan, and Mangarrayi, and languages from the non-Pama-Nyungan families Bunuban, Daly, Garrwan, Gunwinyguan, Jarakan, Maningrida, Mirndi, Nyulnyulan, Rembarrnga-Ngalakgan, Tangkic, Worrorran, and Yangmanic. The items collected for the survey were ordinal and cardinal numerals (as many as were attested in the database), along with quantifiers such as 'some', 'a few', 'many', and 'all'. For each language, materials were then coded for properties relating to the numeral system, as follows:

- The forms for basic numerals; subsequent analysis provided data regarding:
    (i) the extent of the numeral system (that is, how high the numerals go);

(ii) how (and whether) numerals are combined to form higher numerals;
(iii) etymologies of numeral forms (and whether any are likely loans);
(iv) whether there is information about the use of numerals for vague counting.

- Whether there are ancillary counting systems.
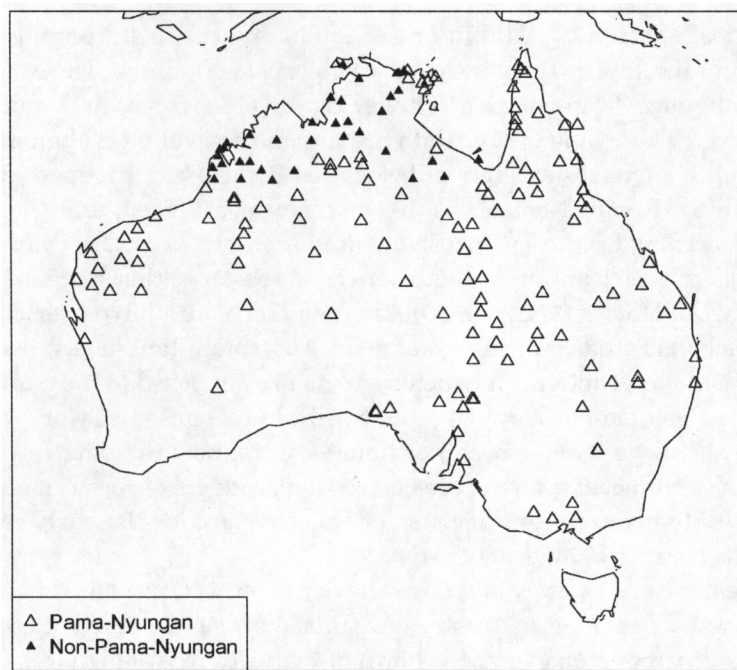
Map 1. Location of varieties in the survey.

We have no data at present on the extent of ordinal numeral systems. Morphologically derived ordinal numerals (as are found, for example, in Dyirbal [Dixon 1972:240–41]) do exist in some languages, but we currently lack the data to be able to assess their frequency across the continent. There are words that translate English ordinals in the database, but they are clearly not part of ordinal systems. For example, terms for 'first' in our lexical database refer to the position of someone in a line (or the position of an item in a temporal sequence), rather than the numeral. Likewise, 'second' usually means 'following' or 'behind' and is not a true numeral (see also Nekes and Worms 2006:124). See section 4.2 for birth order names.

We did not survey grammatical number systems in this work; thus, the extent of singular, dual, paucal, and plural marking in nouns and verbs is not discussed here. The survey materials presented here clearly show that lexical number (i.e., numerals) and grammatical number are not the same in Australian languages overall (pace Hale 1975). There are also differences in the ways in

which grammatical and lexical number marking are used, even in languages with very small numeral systems. For example, to our knowledge dual pronouns do not have vague numeral reference, even in languages where numerals denote inexact quantities. (See further section 3.3 below.)

3. Results.

3.1. Extent of systems. All but one of the languages in the sample showed limited numeral complexity, in keeping with previous findings. The exception is the Kulin language Tyapwurrung, where Dawson (1881:xcvii) lists numerals to 'one hundred'.[6] The Tyapwurrung data are difficult to evaluate. Numeral limits for other Kulin languages are 'ten' or below (see Blake 1998; Hercus 1986) apart from the Kuurn Kopan Noot forms given in Dawson (1881), and the circumstances of recording for the Tyapwurrung data are unknown. The Tyapwurrung numeral system is a clear outlier, but there is no reason at this stage to doubt its authenticity. As Harris (1987:29–30) has noted, scholars have overlooked the evidence for elaborated counting systems in Australian languages, perhaps in part because of a perception that such systems are supposed to have low limits. This point is also made by Dawson (1881:xcviii) in response to Tylor (1871:220) that numerals above 'two', 'three', or 'four' are unknown in Australia. While some extensive numeral systems are clearly the products of recent elaboration, there appears to be no reason to consider this the case for the early reports of numerals discussed by Dawson (1881).[7]

Because almost all survey languages have numeral term limits below 'twenty', they may be classified as "restricted" in the sense of Comrie (2005). Still, upper limits do vary extensively. As shown in figure 1,[8] most languages (n = 139) have an upper limit of 'three' or 'four', but several languages have contiguous systems that extend up to 'five' (n = 21), 'six' (n = 8), or 'ten' (n = 8).[9] According to early sources (Ridley 1875), a variety of Gamilaraay extended up to 'twenty'. These forms do not appear in Ash, Giacon, and Lissarrague's dictionary (2003), however. The word for 'ten' supplied by Ash, Giacon, and Lissarrague (2003: 160), *banay*, is a neologism developed by the Walgett language program; this replaces an earlier recorded word, *bularruu maa* 'two hands'. Two other languages, Gooreng Gooreng (Waka-Kabi) and Tagalag (Paman), have systems extending to 'ten' and 'nineteen', respectively.

There are nine languages where the only numerals recorded were 'one' and 'two'. They are represented in figure 1 as having an upper limit of 'two', but the systems may extend further. In all cases but one, there are reasons to believe that the data might be incomplete. For example, for Yarluyandi and Marrgany, records of the language from the last speakers were made over a short period of time (and it is quite possible that other numerals were simply not recorded). Mpalitjanh data are also scanty. Malngin, Jiwarliny, and Yulparija are all varieties of languages where other dialects have larger numeral systems.

Jiwarliny is the westernmost dialect of Walmajarri, which has numerals up to 'four' (Richards and Hudson 1990); Malngin is a variety of Jaru, which is also recorded as having numerals up to 'four' (Tsunoda 1981). Yulparija is a Western Desert (Wati) language. In that dialect chain, there is some variation in the maximal number of numerals of well-described varieties, with one variety (Kukatja) attesting numerals up to 'ten' and others showing 'three', 'four', or 'five'. We are thus unable to determine whether the omission of 'three' in the Yulparija data represents a real gap or is a data artifact. The final language with only two numerals, Wangkumara, is also likely to have a source problem, since a different source for the language (Robertson 1984) attests *parrkuwatyu* for 'three'.
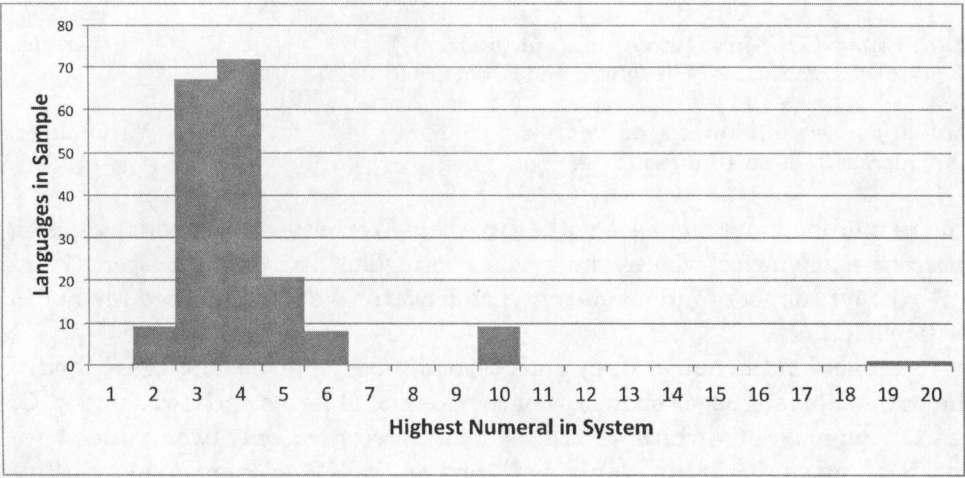

Figure 1. Histogram of upper limits of numeral systems.

In summary, contrary to Hanke's (2005:17) claim that 'three' is found "seldom" in Australian languages, only 5 percent of the languages in our sample have an upper limit of 'two'.[10] The great majority (74 percent) of the languages have an upper limit of 'three' or 'four'. An additional 11 percent have an upper limit of 'five', and the remaining 10 percent have limits that vary from 'six' to 'one hundred'.

3.2. Internal structure of the numeral systems.

3.2.1. Compositionality. Most (56 percent, n = 106) of the surveyed languages combine smaller numerals to create larger ones.[11] None of the languages have compositional forms for 'two', but in the higher numerals, we find considerable variation in whether and how numerals are combined.

Almost all (94 percent, n = 177) of the languages in the sample have terms for 'three.' The majority (n = 117) of the languages with 'three' have an atomic

(i.e., noncompositional) form. Forty-nine languages have a compositional 'three';
in forty-three of these languages, this consists of the form for 'two' plus the form
for 'one' (hereafter, 'two'+'one'), but other possibilities exist—'one'+'two' in
Gooreng Gooreng, 'two'+'one' or 'one'+'two' in Wiradjuri and Ringu Ringu,
'two'+'one' or 'two'+'a few' in Nyikina, and 'two' plus an unidentifiable element
in Wangkumara (Thargominda) and Yarluyandi (North Shore of Lake Eyre).
The preference for atomic 'three' is stronger in languages whose numeral sys-
tems go no higher than 'three'; in other words, if a system has a limit of 'four' or
higher, it is more likely than a system with a limit of 'three' to have a composi-
tional 'three'. There are nine languages that have a compositional form as an
alternative to an atomic 'three'. Both Warlpiri and Warumungu, for example,
have both atomic and compositional 'three', as shown in (1a) and (1b).[12]

(1a) *jirrama-(kari)-jinta* 'three' (lit., 'two (and) one')   (Warlpiri)
     marnkurrpa 'three' (Laughren and Hoogenraad 1996)

(1b) *kujjarra-yarnti* 'three' (lit., 'two one')   (Warumungu)
     yurrkarti 'three' (Simpson 2002)

The remaining two languages with 'three' are Worrorra and Yawijibaya, both
Worrorran languages, where numerals are formed using a quantificational stem
inflected for number (and person and gender); this will be discussed further in
section 3.3.2.
    Atomicity in the numerals extends beyond three for some languages, though
the number of languages showing noncompositional forms declines sharply. Of
the 112 languages for which a word for 'four' is recorded, only twenty-nine have
an atomic term. Sixty-nine languages have a 'two'+'two' compound, and five
languages have a term that is either partially opaque (based on 'two' plus an
unidentifiable element) or an etymological derivative of 'two'+'two'. For ex-
ample, Mudburra has a form *kutyatyarra* for 'four', which is transparently a
contraction of *kutyarra–kutyarra* ('two'+'two'). Five languages have an atomic
term alongside a 'two'+'two' form for 'four'. Walangama uses 'three'+'one' for
'four', and a form based on 'three' is also available in Wakaya, where there is
also a 'two'+'two' form. In Tiwi, *jatapint* 'four' is a compound of 'one' and an
unknown second item.
    There are fifty-four languages in the sample with attested words for 'five'.[13]
Of these, most (n = 36) have atomic terms for 'five', including twenty-three
languages whose 'five' is based on 'hand'. Eleven languages have 'two'+'two'+
'one' or 'two'+'three' (in the latter case, 'three' is atomic) for 'five'. Four lan-
guages have both 'hand'-based and 'two'-based forms. Two varieties of Kaurna
(Thura-Yura) use 'four'+'one' for 'five', where 'four' is noncompositional. Final-
ly, Pitjantjatjara/Yankunytjatjara (Wati) has both a 'two'+'three' form and a
form based on English *five*.

'Six' is found in twenty-one languages surveyed, and its compositionality patterns appear to be dependent on the upper limit of the system. Seven of the twelve languages with a limit of 'ten' or higher use 'five'+'one' for 'six', whereas only one of the eight languages with a limit of 'six' uses 'five'+'one'. Six of these eight languages whose systems end at 'six' have compositional forms for 'six', and they are formed in five different ways: 'two'+'two'+'two', 'three'+'three', 'two'+'three', 'four'+'two', and 'five'+'one'. 'Six' in Ngarluma is atomic, and the Narangga form is derived from a birth order term (see section 4.2).

Among the twelve languages that have numerals higher than 'six', eight languages use compositional forms for the higher numerals. All of these eight use an atomic 'five' as a base (see section 3.2.2), except for Gooreng Gooreng, which combines 'one' and 'two' additively to form the numerals through 'nine'. The other four languages that extend beyond 'six' have atomic higher numerals.

As discussed above, 'three' and 'four' may each be compositional or atomic. 'Three' is more likely to be atomic than compositional; the reverse is true for 'four'. In table 1, we give the combinatory possibilities for the two most common forms of 'three' (atomic and 'two'+'one') and 'four' (atomic and 'two'+'two'). Some languages allow for multiple patterns, so the table gives the number of languages that require each pattern as well as the number of languages in which a given pattern is available (either as the only pattern or one of several patterns).

Table 1. Languages with the Most Common Compositionality Patterns

PATTERN	OBLIGATORY	AVAILABLE
'three', 'two'+'two'	31	41
'two'+'one', 'two'+'two'	27	35
'three', 'four'	25	29
'two'+'one', 'four'	4	7

We can see that among the languages surveyed, there is a strong dispreference for a system that has 'two'+'one' for 'three' and then reverts to an atomic form for 'four'. The most common pattern is 'three', 'two'+'two', an example of which is given in (2).

(2) *tyintaku* 'one'   (Gurindji)
    kutyarra 'two'
    murrkun 'three'
    kutyarra kutyarra 'four' (lit., 'two two')
    kutyarra kutyarra yangi 'five' (lit., 'two two another') (McConvell p.c. 2010)

Eighty-three (44 percent) of the languages in the sample lack compositional numerals entirely. Most of these languages (n = 60) have two or three numerals only, while another fifteen languages have atomic terms up to 'four'. Five languages from different parts of the country (Wirangu, Warriyangka, Kurtjar,

Uradhi, and Yidiny) have noncompositional terms up to 'five'. Tagalag's nineteen terms are all atomic. The numerals in Rembarrnga's system are noncompositional, but they are all words for species of wallaby (see section 4.1).

3.2.2. Bases and series. As discussed by Hanke (2010), most numeral systems in the world's languages are serial, combining numerals to create additive series, multiplicative series, or both. An additive series is built by adding a sequence of numerals to an additive base, and a multiplicative series is built by multiplying a multiplicative base times the numerals in a sequence (Greenberg 1978:258). Most commonly, languages use both types of series in an integrated system; an additive-multiplicative base serves as both an additive and a multiplicative base. These three types of bases may be summarized by Comrie's definition for base: "The value n such that numeral expressions are constructed according to the pattern $xn + y$, i.e. some numeral x multiplied by the base plus some other numeral y" (Comrie 2005:530). An additive-multiplicative base is one in which both x and y are obligatory (but allowing for the possibility of deletion of x when it is 'one' [Hanke 2010:86]). An additive base does not include x, but requires y, and a multiplicative base does not include y, but requires x.

An alternative definition of a base is given by Hammarström:

The number n is a base iff

1. the next higher base (or the end of the normed expressions) is a multiple of n; and

2. a proper majority of the expressions for numbers between n and the next higher base are formed by (a single) addition or subtraction of n or a multiple of n with expressions for numbers smaller than n. [Hammarström 2010:15]

Such a definition is functionally equivalent to Comrie's definition for unbounded systems. For our purposes, however, Comrie's definition of a base is more usable, because it does not refer to a higher base than the one being defined. For most of the languages surveyed here where a base of some kind can be defined, there is a single base. Therefore a definition of a base that makes reference to a higher base is unworkable.[14] As seen below, the definition and identification of a base constitutes the most significant difference between productive and small numeral systems.

Most of the Australian languages surveyed here (n = 106) use some kind of base, whether additive, multiplicative, or additive-multiplicative. In this, they exhibit similarities to unrestricted numeral systems. Almost all of these systems (n = 97) use 'two' as a base, but because a form 'two'+'two' meaning 'four' could be interpreted as either 'two plus two' or 'two times two', it is difficult to determine whether these languages use additive or multiplicative series, except in the cases where a word for 'six' is recorded. The rarity of terms for 'six' of the form 'three'+'two' or 'two'+'three' (interpreted as 'three times two', 'two times three') implies that 'two' is not usually a multiplicative base in these languages;

such a term is, however, found in Koko-Bera (Black n.d.), where *pulʋkonówʋrr* is a compound, 'two'+'three'. Garrwa has *kujarra baki kujarra baki kujarra* 'six' (lit., 'two and two and two'), implying an additive base. If we assume that the sequence of numerals added to an additive base nearly always begins with 'one' (Hanke 2010:68), we would predict that systems with atomic 'three' instead of compositional 'two'+'one' use 'two' as multiplicative rather than additive base, but it is impossible to test this hypothesis without higher numerals. Koko-Bera supports this hypothesis, but the rarity of relevant systems in the data set makes it impossible to test this hypothesis further.

A few languages in the survey (n = 9) use 'five' as a multiplicative base, creating 'ten' from 'two'+'five' (meaning 'two times five'); in these cases, the base is derived from 'hand' even if there exist separate base-2 forms for 'five'.[15] Two languages use 'five' as an additive base; this base is a 'hand' derivative in both Muruwari (Oates 1988) and Wiradjuri (Mitchell 1848). At least two languages use 'four' as an additive base for at least part of the system, which is crosslinguistically rare and previously unattested for Australia (Hammarström 2010:24–27). Muruwari (Oates 1988), for example, has *kapulanta* 'four' and *kapulanta kapulanta* for 'eight' (lit., 'four four').[16] Base 3 is also rare (Hammarström 2010:22–24); it may be present in Adnyamathanha (McEntee and McKenzie 1992), Gamilaraay (Ash, Giacon, and Lissarrague 2003), and Yir Yoront (Alpher 1991). Adnyamathanha has *urlparrina* for 'three' and *urlpa-urlparrina* for 'six', but McEntee and McKenzie (1992) list no forms for 'four' or 'five', so it may be that *urlpa-urlparrina* is not truly part of the numeral system. In Gamilaraay, 'six' may also be expressed by the more productive base-5 additive pattern, so the reduplicative base-3 form may be a minor variant. Yir Yoront has two words for 'six': *wapayrr wapayrr* ('three'+'three') and *pulkoyrr*, a compound of *pul*, the third person dual pronoun, and *koyrr* 'two'.

Hanke (2010:70–71) argues that base-2 systems are nonserial (calling them "pair sums") because the asymmetry between the additive base and its addends is less obvious than with higher bases. In our survey, there are base-2 systems (n = 4) that have atomic 'three' and use it as an addend along with 'two' and 'one'. Such systems cannot be analyzed as pair sums and must be treated as serial patterns built on a base 'two'. If the pair sum hypothesis is to be maintained, the existence of these truly serial base-2 systems forces us to pursue a bifurcated analysis for base-2 systems, where systems are treated differently on the basis of their addends. This seems like an artificial dichotomy, so we reject the pair sum hypothesis and treat all base-2 systems on par with series with higher bases.

Hanke has claimed that "in principle, the minimal range of a[n additive] series is just two adjacent numerals, but in practice the smallest series consist of 5, 10, or multiples of these numbers" (2010:68). In our survey, the most common series range[17] is two (n = 54), but three is also quite prevalent (n = 37). Five languages have a series range of four, and four languages have a series range of

five. Gooreng Gooreng has a series range of eight. Thus, we have a preponderance of evidence for minimal series ranges below five.

In summary, many Australian languages combine numerals to create higher ones, and these languages vary especially with respect to whether the term for 'three' is compositional or atomic. Atoms for terms above 'three' are rare: only thirty-four languages in the survey have an atomic form for 'four', and five of these languages also have a compositional form for 'four' based on 'two'. When we analyze the surveyed systems in terms of bases and series, we see that most of the languages are base 2, but there are a few possible base-3, base-4, and base-5 systems as well. Hanke's (2010) treatment of base-2 systems as a separate "pair sums" category appears to be unjustified, and there is evidence for minimal series ranges below five.

3.3. Numerals within the grammatical system.

3.3.1. Vagueness of numeral denotation.   According to Dixon (1980:107–8), following Hale (1975), numerals do not comprise a separate word class in most Australian languages.[18] Rather, words that translate English numerals are, in fact, adjectives or quantifiers by distributional or morphological criteria. (Figures are not given by either author for the number of languages that do or do not have separate numeral classes.) Hale suggests that Australian numerals are a type of indefinite determiner used for vague qualification. For example, in Dyirbal (Dixon 1972:240), *mundi* means 'some', and stands in sequence following the numerals 'one', 'two', and 'three'; in Jaru (Tsunoda 1981), *murrkun* translates 'three', but it could also be used for 'several' items. The same holds for Warlpiri. Dixon (1980) appears to generalize from Hale's comments on Warlpiri to regard "almost all" Australian languages as fitting this profile (he does not say which languages might be exceptions). Hale also notes that a numeral system of the form 'one', 'two', 'paucal' (= 'some'), and 'plural' (= 'many') is parallel to common grammatical number systems, as are found, for example, in nominal and pronominal number marking in languages elsewhere in the world.

This view is not supported in the aggregate, though we do not contradict it for the specific languages for which it has been claimed. It is clearly the case that some Australian languages show marginal numerals that have vague reference and are frequently used to describe inexact quantity. Equally clearly, however, other languages do not have numerals that are used in this way. Australian languages vary in the extent to which numerals constitute a distinct word class, and in whether they allow vague readings (or denote exact quantities). We also find analytical problems in treating vague numerals as equivalent to grammatical number agreement, as discussed in section 3.3.2.

Positive information regarding vagueness was available for one hundred (roughly half) of the languages in the survey.[19] The languages in this survey vary in the extent to which the numerals allow a secondary meaning of 'some',

'several', or 'a few' in addition to their use as numerals. About half of these languages (n = 41) allow vague readings for at least some numerals, while the rest (n = 59) do not, having distinct vague quantifiers and information that the numerals are not used with vague reference. These findings support McGregor's (2004:206–7) findings for the Western Kimberley region of northern Australia, where languages differ in concreteness of numeral reference across the area.
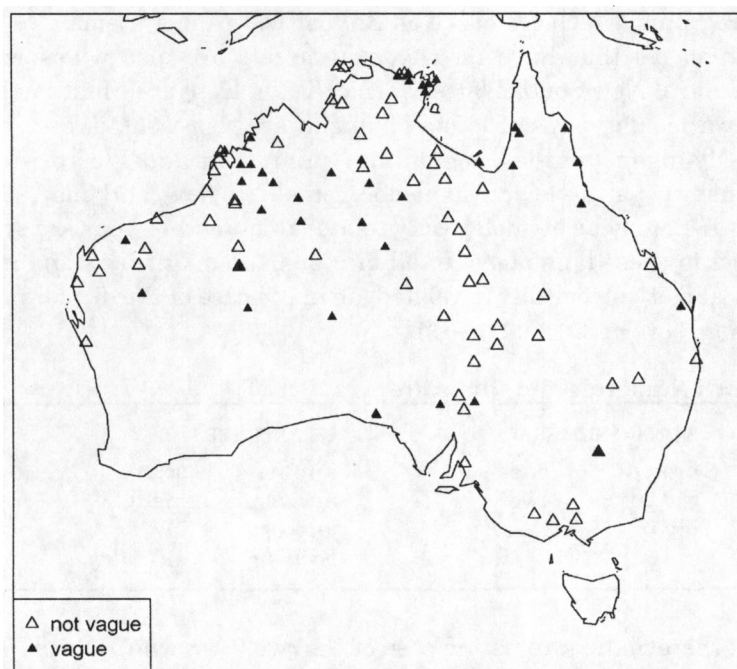
Map 2. Languages with vague and nonvague numerals.

The languages that allow vague numerals include Bunuba and Gooniyandi (Bunuban), the Maningrida family, and, within Pama-Nyungan, a number of Ngumpin-Yapa and Wati languages, Thura-Yura languages, and Yolŋu languages (see map 2). Some Eastern Nyulnyulan languages also allow vague numeral reference (though the Western Nyulnyulan language Bardi [Bowern 2012a] does not). Only one language (Warlmanpa) is recorded as allowing indefinite reference for the numeral 'one',[20] and no language has indefinite quantification for the numeral 'two'.[21] In many Western Desert varieties, the word *marnkurrpa* 'three' can also be used to mean 'several' or 'a few'; the same is true for the widespread Yolŋu word ḻurrkun' (examples are from Zorc [1986] for Dhuwal, Dhuwala, Dhaŋu, and Ritharrngu; Waters [1989] for Djinang; and Bowern's Yan-nhaŋu field notes). The vague readings of 'three' vary from 'few' (in Ngarluma [Hale 1988]) to 'several' (Koko-Bera [Black n.d.]) to 'many' (Warungu [Tsunoda 1972]; Yidiny [Dixon 1991]).

The Western Desert variety Kukatja (Peile 1993) appears to have a distinction where count nouns may receive a nonvague reading for *murntu* 'four', but mass nouns receive the interpretation 'much'. One language, Wiradjuri, is ambiguous, with the older sources implying that *madu* 'five' can also be used for either 'four' or 'many', while Hosking and McNicol (1993) provide additional quantifiers and imply that numerals are not vague. This may represent a change in the language or a reinterpretation of numerals in line with school teaching programs, which are based on English counting systems.

The languages that have no vague numerals are also widespread. The Marrngu, Central New South Wales, Karnic, Kulin, Kanyara-Mantharta, Kartu, and Warluwaric subgroups of Pama-Nyungan are represented, as are several non-Pama-Nyungan families, including Nyulnyulan and Worrorran. These languages have separate lexical items for 'some', '(a) few', and 'many'. Though these quantifiers may be etymologically related to numerals, they are synchronically distinct in these languages. In other cases, the quantifiers and numerals are distinct and etymologically unrelated, as in the case of Bardi, shown in table 2 (see further Bowern 2012a:279–83).[22]

Table 2. Bardi Numerals and Quantifiers

ATOMIC NUMERALS		QUANTIFIERS	
arinyji	'one'	*aranga*	'some'
gooyarra	'two'	*niimana*	'many'
irrjar	'three'	*jalboorr*	'few'
		boonyja	'all, every'

Finally, there is the interesting case of Garrwa (Garrwan), which has vague terms based on, but distinct from, its numerals. For example, 'three' is *kujarra yalku* ('two'+'one') and 'four' is *kujarra kujarra* ('two'+'two'), while 'a few' is *kujajarra* (Furby and Furby 1977).[23] Thus, in some languages there are etymological relationships between the numerals and quantifiers, though they may be distinct synchronically. We have not found this to be a widespread pattern, however.

Thus, vague numerals are not a general and necessary property of small numeral systems; rather, it is a point of variation that should be treated separately from numeral system size.[24] One might wonder whether the impact of European colonization on language use in Aboriginal communities, particularly the use of Aboriginal languages in counting in mission schools, has led speakers to suppress vague usage of numerals. This may be true in some cases, but it is difficult to see why this would be the case, and what would cause languages in some areas to show such a change, but not others.

3.3.2. Numerals and quantification. While we lack detailed crosslinguistic information for the ways in which the numeral systems interact with grammatical marking, we do have some evidence that numerals are not felicitously

analyzed as indefinite determiners for all languages, and that the numerals variably comprise a distinct system. We have no reason to doubt Hale's (1975) analysis of Warlpiri, for example. For Worrorra, Clendon (2000:335–36) shows that numerals are formed by adding grammatical number suffixes to a quantificational stem =*yarrungu*, which inflects for both person and number, as shown in (3a)–(3c). When declined without a numeral suffix, =*yarrungu* means 'many' (cf. *ajarrunguya* 'many of them'); here it is glossed as QUANT (for "quantification marker"), following Clendon (2000). Forms are quoted in the third person feminine.

(3a) *ninjarrungunya*   (Worrorra)
    nyiN=yarrungu-nya
    3FEM=QUANT-FEM
    'one [feminine]' (Clendon 2000:336)

(3b) *ninjarrungaarndinya*   (Worrorra)
    nyiN=yarrungu-ᵏwarndu-nya
    3FEM=QUANT-DUAL-FEM
    'two [feminine]' (Clendon 2000:336)

(3c) *ninjarrungoorrinya*   (Worrorra)
    nyiN=yarrungu-oorri-nya
    3FEM=QUANT-TRIAL-FEM
    'three [feminine]' (Clendon 2000:336)

In Worrorra, the same suffixes (singular null, dual -ᵏ*warndu*, trial -*oorri*, and plural -ᵏ*wurri*), are used in numeral formation, pronominal number marking, and verbal argument agreement. The Worrorra forms do not have vague numeral reference (Clendon 2000:335). This language is the only one, to our knowledge, in which lexical numerals and grammatical number marking match so closely. Even other members of the same family show different patterns. Unggumi, for example, has *wintharringga* for 'one', *wintharringarndoo* for 'two', but *ngarlgurru* for 'three'. The first two numerals are presumably cognate with the Worrorran forms for 'one' and 'two', but 'three' is not. Ngarinyin has atomic 'one' and 'two', but compositional 'three' ('two'+'one') and 'four' ('two'+ 'two').

A crucial difference between Worrorran and many other Australian languages is the cooccurrence of determiners and numerals, implying that the two belong to separate classes. In many of the languages in which the word class of numerals is discussed, numerals behave distributionally like adjectives and may occur with demonstratives, as in Warungu in (4).

(4) *Ngalnga* *yinda*   *bindaynga-lgo* *yino*   *gornggal-∅*   (Warungu)
    PROH   2SG.ERG let.go-PURP   2SG.GEN husband-ACC

    ngona-∅ ***nyon.gol-∅*** *jarribara-∅*
    that-ACC   one-ACC   good-ACC
    'Don't let that one good husband of yours go.' (Tsunoda 2011:347)

Parallels in form between grammatical number marking and lexical numbers are found in duals, but not in other parts of the system. For example, cognates of the lexical numeral 'two', *kutharra* or *kutyarra*, are found as dual number markers in languages from several subgroups of Pama-Nyungan (Dixon [2002:116–17] gives examples from Gurindji and Ngarluma). A dual pronoun *pula* is also widespread (see section 3.4.2) and is also found as both a numeral and a dual marker. We do not find equivalent parallels with either 'one' or 'three', however. Thus, while there is a formal link between grammatical dual marking and the numeral 'two', this does not extend to other parts of the numeral system, and therefore makes the equation of lexical and grammatical number marking tenuous. After all, numerals are a source of grammatical number marking in languages of the world without restricted numeral systems (Corbett 2000:135).

We should note that there is some inconsistency in the more general argument that numerals are vague quantifiers or indefinite determiners. Linking the verb agreement to the numeral system has the implication that the verb agreement is also vague for number. This is not true, at least for some languages. Since there is variation in whether the numeral systems exhibit vague reference, it is possible also that the grammatical number systems vary as well. We have no continent-wide information on this point, but we do have information for some languages. In the Nyulnyulan language Nyikina (Stokes 1982), for example, the numerals above 'two' are vague, but there is a "unit augment" agreement system (which is not vague for three referents) in verbal marking. In the Nyikina pronominal system, singular reference pronouns may take a suffix *-mirri*, which adds a single third person referent to the group denoted by the pronoun. Thus, *yayu* is the first person dual inclusive (first person plus second person) pronoun, and *yayumirri* is a pronoun that refers to first person, second person singular, and a single third person referent.

3.4. Etymology of numerals.   While there is not space to discuss detailed etymologies of all forms, some trends are noticeable in the data. Since numerals derived from smaller numerals are surveyed in section 3.2 above, only the etymologies of atomic forms are considered in the present section. Most noticeable is the difficulty in reconstructing numeral terms to any time depth, either in Pama-Nyungan or in other Australian families. This might suggest that either numeral concepts are relatively recent in Australian languages, or (more likely) that small numeral systems are relatively unstable.

We note that numerals are not universally stable; there is extensive variation in the world's languages on this point, both as to which families show atomic numeral stability, and as to which numerals are stable in those families. In Austronesian, for example, lower numerals are quite stable, while higher numerals show much more variation, due no doubt in part to the plethora of bases and subbases found in different languages in the family, along with the existence of both additive and subtractive bases in different regions.[25] Within

Indo-European, numerals below 'ten' are highly stable, though higher numerals show evidence of different bases in various subgroups (Beekes 1995; Fortson 2004). In Algonquian, while numerals below 'five' are stable and easily reconstructible, those between 'five' and 'ten' show numerous changes (Rhodes and Costa 2003); the same is true for Bantu (Meeussen 1967; Zerbian and Krifka 2008).

3.4.1. Forms for 'one'.   No form for 'one' is reconstructible uncontroversially to Proto-Pama-Nyungan. Languages of the Yolŋu, Mayi, and Maric subgroups point to *waŋka-, but this nasal+stop cluster in Yolŋu should give reflexes of the form waŋa (homorganic nasal+stop clusters are reduced; cf. *waŋka 'say' > waŋa). The correspondence of Yolŋu *waŋka to an identical form in Mayi and Maric is thus irregular, and points perhaps to a loaned form in Yolŋu. A different form, *kurnu, is found in a number of languages in the Karnic and Yardli groups. In the southwest of the country, in Nyungic the form *kayanu is reconstructible, and several other forms appear in particular subgroups within Pama-Nyungan as well, shown in table 3.

Table 3. Additional Pama-Nyngan Forms for 'One'

*wanytyu	found in Marrngu and wandhu in Buwandik (though not in other Kulin languages); this is likely simply coincidence, given their wide geographic separation
*kunytyi	in the Ngayarta subgroup
*kurrika	in Kanyara-Mantharta and Kartu (kunytyi in Ngayarta is then presumably a replacement of the earlier term)
*kutyu	in Wati (cf. also kutyupa 'another', kutyuŋka 'the same', Pitjantjatjara kutyuli 'everyone', Kukatya kutyutya 'unmarried')
tyinta	widespread in the Ngumpin-Yapa group, but probably diffused, as it does not seem to fit reconstructed sound changes (McConvell and Laughren 2004)
*kuma	in Thura-Yura (Simpson and Hercus [2004] reconstruct *kutyu, however, but that would appear to be a loan from Wati)
*mithuŋ	in Yuin languages
*waakal	in Kuri
*yarrkulu	in Warluwaric
*NanyantV	in Arandic; Warumungu yarnti might be a loan from Arandic
*watyin	in Maric (Eastern Karnic also shows watyu, which is probably a loan)
*kaypa	in Kulin
*yaparu	in Bandjalangic

None of these words are traceable to other words with other etymologies thus far, and they are not relatable to each other.[26] McGregor (2004:207) gives a schematic representation (semantic map) of differing senses of words for 'one' in

Australian languages, which include reference related functions (both indefinite and definite ['the particular one']), temporal functions ('one by one, one at a time'), and quantificational functions ('one each' and 'small size'). The form *kutyu* is found in a number of other languages in the northwestern region of the Pama-Nyungan family in derived forms (e.g., Northern Nyangumarta *kutyurl* 'alone'); similar semantic extensions are found with some of the other numeral words, where, for example, Yulparija *tyintaŋa* means 'alone'.

In non-Pama-Nyungan families, the Nyulnyulan languages have *warinyji* uncontroversially. Worrorran languages form numerals by adding grammatical number inflections to a demonstrative stem. This type of numeral marking is not found elsewhere on the continent to our knowledge.[27] Bunuba and Gooniyandi have *yuwarna* and *yuwarni*, respectively; these are the only two languages in that family.

3.4.2. Forms for 'two'. Unlike in the case of the words for 'one', there are a few widespread forms for 'two'. The most widespread is a form of *kutyarra*, which occurs in fifty-one of the languages in the sample and is found all over the country.[28] This form shows evidence of diffusion; it is sufficiently widespread in similar forms that it is unlikely to be original in all the languages (though there is also evidence that it is reconstructible fairly far back, since it undergoes regular sound changes in the languages of Cape York Peninsula, for example). It is loaned into Nyulnyulan languages (Bardi *gooyarra*, Nyikina *kujarra*, where it is reconstructible as a loan into Proto-Nyulnyulan as *kutyarra*). It is also a loan in both Wambaya and Jingulu.

A few languages have a form based on *mandu* 'pair'. The languages with this form as a numeral form an areal group from three different subgroups (Diyari and Wangkangurru in Karnic, Wadikali in Yardli, and Guyani in Thura-Yura, all spoken in the region of Lake Eyre in central Australia).

The other main reconstructible form is *pula*, which is also widespread as a third person dual pronoun. There are also words for 'two' in a number of different subgroups that augment a stem *pula*– with something else, such as a plural marker or what appears to be a reduced plural marker.[29] These different forms are scattered across the country, as shown in table 4, and have the pattern of likely parallel innovations (i.e., they do not have either a clear subgroup-based or a clear areal distribution).

Table 4. Geographical Distribution of Forms for 'Two' Based on *pula*

pula.rri	in a number of languages, especially in the eastern part of the country, including Gumbaynggirr, the Yuin-Kuri language Wodi-Wodi, several languages of Central New South Wales, Proto-Maric, and Bandjalangic
pula.karra	in the Mayi subgroup and the Yuin-Kuri language Awabakal
pula.tya	in the Kulin subgroup

3.4.3. Forms for 'three'.

Forms for the word for 'three' are also varied and difficult to reconstruct across the continent. The common forms in the data are given in table 5. In addition to the atomic forms, there are also compositional terms of the form 'one'+'two' or 'two'+'one' (for which see section 3.2.1 above). The only one of these forms that has any potential cognates outside the domain of numerals is *kulpari* in Karnic.

Table 5. Forms for 'Three'

tyarrku	in Kanyara-Mantharta and Ngayarta languages
murrku(r)na	in Ngumpin-Yapa, though it shows signs of borrowing, since it is also found in non-Pama-Nyungan languages in the area (Wardaman *murrkun* 'three', Miriwoong *merrgen* 'three', Mayi-Kutuna *murrku* 'many')
marnkurr(pa)	in Wati (and borrowed, presumably, into Kaurna, though with a different final syllable)[†]
ḻurrkun'	in Yolŋu, of unknown origin
kurrpara	in Mayi, also in some nearby Maric languages, including Wadjabangayi (⟨kurbaddie⟩) and Yirandhali, and the Karnic language Pirriya (*kurrpara*)
kulpari	in Karnic; this form is presumably cognate with *kurrpara*, though the details are unclear (most Karnic languages have a 'two'+'one' form for 'three'); Luise Hercus (p.c. 2010) suggests that this is related to *kulpari* 'emu', the etymological connection being that that emus have three toes
kulipaa	in Central New South Wales

[†] Kaurna has *marnkutye*. In Western Desert, *-pa* is regularly added to words that end in a consonant; the languages have word structure constraints that require all prosodic words to end in a vowel.

3.4.4. Summary and other forms.

Numerals for 'one' and 'two' are atomic in all languages in the sample. The latter is more uniform than the former, though reconstruction to Proto-Pama-Nyungan is difficult. Most likely is a form *kutyarra* for the numeral 'two' and a dual demonstrative *pula*.

There are forms in the data that are borrowed from English, though not for the most basic numerals. Loans from English are sporadically recorded in the lexica for the languages in the sample. For example, a loan from English *six* is recorded in both Nyangumarta and Martu Wangka, but in neither case are numerals for 'four' and 'five' listed in the sources. Martu Wangka's numeral 'six' is *jikaji* from English *six card*, the six in a deck of cards.

There are also a few cases where numerals change into other numerals. In Diyari (Austin 1981), the word for 'three', *parrkulu*, is cognate with the word that means 'two' in other Karnic languages. The word for 'two' is replaced by a word reconstructible in the meaning 'pair'. Since in most of Karnic 'three' has the form 'two'+'one', we assume that the compound numeral lost its second

member (i.e., this is not a case of vague numerals). Such a change is well attest-
ed in compounds in other languages.[30]

Twenty-eight of the languages use a form meaning 'hand' to denote 'five' or
'ten' ('five' is *ŋanuŋalja* 'my-wrist' in Worrorra). Examples include *mara* in
Wangkumara and Yandruwandha (from Proto-Pama-Nyungan *mara* 'hand')
and *goŋ* 'hand' in Yolŋu languages. This polysemy is also found in the Yolŋu lan-
guages that have replaced the Proto-Yolŋu word (e.g., Yan-nhaŋu *goku–waḻip*
'five' [lit., 'hand-one']). 'Foot' is used for numerals between 'ten' and 'twenty' in
languages such as Gamilaraay, where *balar-ya dinna* is 'twenty' (lit., 'two-ya
foot') and *balar-ya marra* is 'ten' (lit., 'two-ya hand'); the meaning of the suffix
-ya is not recoverable. There are few additional etymological sources for numer-
als—a contrast to hunter-gatherer numeral systems in Amazonia (Epps 2006),
where sources include words for 'twins', 'siblings' or (particularly) 'brothers',
and both 'friend' and 'enemy'. Some recently coined systems derive numeral
names from the shape of Arabic numerals, as discussed in section 4.4 below.
Finally, the Arnhem Land language Rembarrnga (Saulwick 2003) has numerals
based on names for wallaby species; these are discussed further in section 4.1.
As noted above, we find some evidence for the borrowing of numeral terms
within Australia. The Nyulnyulan languages, for example, have borrowed the
numeral *kutyarra* 'two' from their Pama-Nyungan neighbors. There is also
evidence of contact with English, which is discussed below. The evidence for pre-
European contact influencing numeral systems (e.g., in the base-5 systems of
Arnhem Land) is reviewed by Harris (1982). While there is substantial evidence
for culture contact between traders from both Makassar and Papua (see, e.g.,
Evans 1992), evidence of its effect on numeral systems is slight. Harris provides
some evidence that the base-5 system in the Yolŋu language Gumatj is a spon-
taneous elaboration based on an earlier system of counting and bundling turtle
eggs. While this could be due to contact, there is also evidence that such systems
are prone to spontaneous elaboration (Evans 2009). Harris compares the Anin-
dilyakwa numeral terms with those in Makassarese and does not find any for-
mal resemblances, a judgment with which we concur. Moreover, extended sys-
tems reported for other Australian languages (Tyapwurrung [Kulin], Gamil-
araay [Central NSW], and Gooreng Gooreng [Waka-Kabi]) are geographically
widely separated and thus clearly not the result of Makassan influence.

4. Auxiliary and secondary counting systems. While the primary purpose
of our survey is to investigate the properties of numeral systems as defined
by Hammarström (2010:11–13), we have found several cases of auxiliary or
secondary number-based systems throughout Australian languages. Following
Hammarström's criteria, these systems are *"spoken, normed expressions* that
are used to denote" a numerical value, but do not apply to "an *open class of
objects* in an *open class of social situations* with *the whole speech community* in
question" (2010:11; emphasis in original). They are thus not numeral systems in

the strict sense, but we survey them here because they are relevant for a wider consideration of "counting" in Australian languages. We also include them because they contrast in properties with the strictly numerical systems we surveyed above.

4.1. Body-tallying. In the Kulin subgroup of Pama-Nyungan, a sequence of body parts from the finger to the head was used for relaying lengths of time in a message delivered by a messenger (e.g., how many days until the meeting to which the addressee is invited) (Howitt 1904:697–703). Body-tallying was also used in the Torres Strait (Ray and Haddon 1891:547–48). As pointed out by Hammarström (2010:12–13), body-tallying is not a true numeral system because it only applies in a specific social context. In Tyapwurrung, however, body-tallying was combined with numerals (Dawson 1881:xcvii).

Rembarrnga (Saulwick 2003) has numerals between 'two' and 'five' that are counted on the fingers (from little finger to thumb) and have the names of wallaby species (e.g., *ngabutj* 'female black red-eyed rock wallaroo' [also 'two'], *garndalppurru* 'female antilopine wallaby' [also 'three']). There is unfortunately no information about how these numerals are used in either the grammar or dictionary. None of the other languages in this region have similar terms that we have been able to discover.

4.2. Birth order names. Ten languages in the survey have a system of names for children based on birth order. Six are from the Thura-Yura subgroup of Pama-Nyungan, described by Schebeck (1973:27).[31] These systems extend at least to 'eight' and as high as 'ten'. Simpson and Hercus (2004:582) reconstruct the first five of these terms to Proto-Thura-Yura. What may be surprising is that these birth-order naming systems may have a higher upper limit than the numeral systems. However, these terms indicate place in a sequence while numerals denote quantity, so we may expect them to have fundamentally different properties. As briefly discussed above in section 2, we found no evidence for ordinal numerals in the dataset, though the relevant information may be missing due to our focus on lexical terms for numerals rather than derivational morphology. Birth order names are also found in Arabana-Wangkangurru (Hercus 1994). Three Yolŋu languages also show a limited (nonexhaustive) system of birth order naming, including 'oldest child', 'second-born child', and 'youngest child'.

4.3. Cattle counting. Gurindji has a series of fifty etymologically opaque terms that were used to count cattle (and earlier, people or game in hunting); see McGregor (2004:205) and Harris (1982:169). The numerals used in this system are not related to the regular numerals in the language, which go up to 'five' (see example (2) in section 3.2.1). Some forms may have cognates in other languages; for example, *panturru* 'one' in the Gurindji counting system is re-

lated to Arabana (and areally in Karnic) *panturu*[32] 'eldest child', and *ŋamirri* 'ten' is a Wati word for marks or notches on a spear. Most of the terms seem to be unique, however.

4.4. Playing card terms and numeral names based on the shape of written numbers.   Several languages have systems for counting playing cards, often using terms that describe the appearance of the Arabic numerals. These terms have also made their way into the language for numerals more generally, particularly through their use in school language programs. The Wati language Kukatja, for example, has a term *wirlki* 'seven' named after a boomerang that is shaped like the numeral. The English name for this boomerang is *number seven boomerang*, which has no doubt facilitated the use of the object name as a term for the numeral. In Warlpiri, other recent numeral coinages are also based on words for items that resemble the shape of the numerals. 'Nine', for example, is *kartaku*, from 'cup' or 'billycan' (which resembles the figure nine when viewed from above), and 'eight' is *milpa* 'eyes' or *mapurlu*, from English *marble*.

4.5. English borrowings for currency.   Some languages, for example the Yolŋu languages in Arnhem Land, use terms based on English words for money in counting currency and other financial transactions. These include *dudala* 'two dollars' and *dindala* 'ten dollars', as well as *tjilba* 'small change' (from *silver*). English or Yolŋu numerals are used for other amounts.

5. Conclusion.   This survey utilizes previously untapped data to provide a better picture of the considerable diversity present even in small numeral systems. The survey results reveal that numeral systems across Australia, while almost universally small, vary considerably with respect to their internal structure, ability to refer to inexact quantities, relationship to grammatical number, and etymological history.

This variation is principled, however. A majority of the languages surveyed combine smaller numerals to create larger ones, almost always creating a base-2 system. In some base-2 systems, the word for 'three' is atomic, while in others, it is compositional. The preference for 'two' as a base across the continent stands in stark contrast to the overwhelming crosslinguistic tendency for numeral systems to be decimal (Comrie 2005). Given the smallness of Australian systems, this is unsurprising, however, since systems with an upper limit below 'ten' cannot be expected to use 'ten' as a base.

While some languages in the sample allow some numerals (usually 'three', but sometimes 'four') to refer to inexact quantities, this is by no means true for all Australian languages, contra Dixon (1980). Similarly, the relationship between numeral systems and grammatical number is rarely as direct as it has been claimed to be.

Variation in small systems does not, of course, provide an argument against structuration (Evans 2003), but it does allow for more refined arguments about the role of cultural practice in language structure. For example, it is fairly frequently argued that languages with small numeral systems have vague reference because of a cultural preference for not specifying inexact quantities (or conversely, a lack of need for precision in higher numerals; see Harris 1982). The prevalence of small systems without vague reference suggests that we should be careful about such generalizations.

Australian numerals on the whole are difficult to reconstruct to any time depth; our most likely candidates for Proto-Pama-Nyungan reconstructions are *kutyarra 'two' and a dual demonstrative *pula. As is found throughout the world's languages, 'hand' is a common source for 'five', but occasionally other salient items may be used as proxies for numerals (e.g., Karnic *kulpari for 'three' based on the emu's distinctive three-toed footprint).

The diversity found within Australian numeral systems suggests that these systems have an important place in numeral typologies. We show in the present article how traditional lines of inquiry such as the nature of series building using bases may be pursued with respect to these small systems. However, the properties of these systems also lead us to raise new typological questions, such as whether numerals may have vague reference and whether vague reference is correlated with system size, which deserve to be investigated across a wider variety of languages.

Notes

Abbreviations. The following abbreviations appear in grammatical glosses: 2 = second person; 3 = third person; ACC = accusative; ERG = ergative; FEM = feminine; GEN = genitive; PROH = prohibitive; PURP = purposive; QUANT = quantification marker; SG = singular.

Transcription. Transcriptions in examples are those of the sources. Common conventions in practical orthographies for Australian languages are to represent dental consonants as ⟨th⟩ (or ⟨dh⟩), ⟨nh⟩, ⟨lh⟩; retroflex consonants as ⟨rt⟩ (or ⟨rd⟩), ⟨rn⟩, ⟨rl⟩; a trill rhotic as ⟨rr⟩; an approximant rhotic as ⟨r⟩; palatal consonants as ⟨ty⟩ (or ⟨j⟩), ⟨ny⟩, ⟨ly⟩; and the velar nasal as ⟨ng⟩.

1. Though see Epps et al. (2012) for a recent survey.

2. "Die Auslassung der australischen Sprachen erklärt sich dadurch, dass für sie, abgesehen von Entlehnungen, fast nur Systeme mit minimaler Komplexität bekannt sind: nach 1 und 2 folgen Quantifikatoren für 'einige' und 'viele', selten gibt es 3 (vgl. Lynch 1998:249)."

3. The relevant reference is Hale (1975). Hale's own description applies specifically to Warlpiri, though he also says that

> it [the Warlpiri counting system] is not a system of numerals, contrary to what one might be led to believe from the literature on Australia, which sometimes identifies languages as having the 'numerals' *one, two, three,* and *many.* The fact is, the indefinite determiner paradigm, as a whole, is not used in counting in Walbiri, any more than are the various definite determiner paradigms. What is true of Walbiri in this regard is, so far as I can tell, true of the other Australian languages which I have any knowledge of. [1975:295]

4. Database development is still in progress; its aim is a collection of all available lexical materials for all extant Pama-Nyungan languages. At the time of writing it contains approximately seven hundred thousand lexical items from 385 languages and about eight hundred doculects (i.e., language varieties as documented in a single source). The data coverage is extensive in some parts of the country (particularly the northwest of the country and certain subgroups such as Central New South Wales), but is patchy in other areas, since data entry is still in progress. This patchiness in data available at the time the research was completed explains the relative undersampling of some languages, including those from the Paman subgroup (Cape York Peninsula). Other areas were undersampled because of difficulties in obtaining sufficient data about numerals. For further information about the database, see Bowern and Atkinson (2012) and Bowern (2012b).

5. They are Arandic, Bandjalangic, Central NSW, Dyirbalic, Gumbaynggiric, Kalkatungic, Kanyara-Mantharta, Karnic, Kartu, Kulin, Maric, Marrngu, Mayi, Ngumpin-Yapa, Ngayarta, Paman, Thura-Yura, Warumungic, Warluwarric, Wati, and Yolŋu. (A full list of Australian languages can be found online: austlang.aiatsis.gov.au.)

6. Dawson (1881) also gives data from Kuurn Kopan Noot, which is closely related.

7. Harris (1982) surveys reports in the literature of systems with numerals above 'five'. He includes Anindilyakwa, which we did not include in our survey due to lack of access to the relevant primary data. Harris also suggests that the Yolŋu language Gumatj shows a productive base-5 system (where 'five' is *rulu* 'bundle', used in counting turtle [*Chelonia mydas*] eggs). Gumatj was not in our sample dataset, though several other Yolŋu languages were, and we found no evidence for similar systems in the Yolŋu lexical data we examined. A third language with a complex numeral system is Tiwi (McRoberts 1990) where numerals are combined with a body-tally system. Our Tiwi data from Osborne (1974) show a smaller system with a limit of 'ten'.

8. Tyapwurrung is not included on the histogram in figure 1 so as to render the values for the lower limit systems more legible.

9. As discussed in section 3.4.4, several languages have terms for 'one', 'two', and 'three', followed by a 'hand'-based term for 'five', 'ten', or both or by an English loan for a single higher numeral. These languages have been coded as having an upper limit of 'three' because the numeral system does not appear to extend higher. See Epps et al. (2012:49–50) for further discussion of this coding principle.

10. But see section 3.3 for discussion of the possibility that some higher numerals have vague denotations, are quantifiers of a more general sort, or both.

11. Hale (1975:295) draws a distinction between true numeral systems (the equivalent of Hanke's productive systems) and systems of exact enumeration (or numeral elaboration), which speakers employ when precision in counting is required, but which are not conventionalized. We have no data on the reality of this distinction for the languages under study, though we do not suggest that Hale is incorrect for Warlpiri. Whether this is true or not, however, we do find some indication that the numeral

combinations are conventionalized. For example, in some languages they show cluster reduction and haplology. In others, they show recursion in compounding, even though such extensive compounding is rare or nonexistent outside the numeral domain.

12. In Warlpiri, the compositional numeral is not frequently used, being associated more with school than with everyday language use (David Nash p.c. 2011).

13. There are more languages with terms for 'five' than languages with limits of 'five' or more because of the coding principle mentioned in note 9.

14. For further discussion of bases in small numeral systems, see Epps et al. 2012.

15. We do not consider the multiplicative base to be 'two' in such languages, because in the sample, the even numerals between 'four' and 'ten' are not formed by multiples of 'two'. This implies that 'two' is not a multiplicative base here.

16. The word *kapulanta* is an opaque derivative of *kapu* 'two'.

17. We could define this more precisely as the cardinality of the set containing a (multiplicative-)additive base and its addends. For example, the series range of Ngarinyin's numerals is three because three numerals ('two', 'two'+'one', and 'two'+'two') are formed from base 'two'.

18. See also Blake (1981:3–4, 33) and Yallop (1982:145). Individual grammars vary on the way they describe numerals; some treat them as adjectives (Oates 1988), others as a distinct type of noun or nominal. In Arabana-Wangkangurru (Hercus 1994:103–5), for example, numerals have the distribution of adjectives, but morphologically they fall into the same declension class as nouns.

19. Evidence for whether numerals were used with vague meanings was taken from field notes and grammatical descriptions, and, in some cases, was inferred from the presence of other vague quantifiers. It is therefore difficult to be definitive in all cases.

20. Languages in the sample differ as to how they mark definiteness within noun phrases. Most commonly, bare nouns are ambiguous, either indefinite or definite; definiteness can be overtly marked, and both indefinite and given (backgrounded) referents are often omitted entirely; see Croft (2007) for an example from Wardaman. We have not yet surveyed such strategies systematically, however.

21. A possible exception to this is Gooniyandi *garndiwa, garndiwangoorroo* 'many' (McGregor 1990), which looks like it is related to the word *garndiwirri* 'two'.

22. Both *arinyji* and *aranga* are probably etymologically based on a root *ara*, which in Bardi is used in a construction meaning 'the one . . . the other . . .', but the other quantifiers are distinct from the numeral system. Moreover, the morphology on *arinyji* and *aranga* is not productive. Speakers of Bardi have been asked explicitly whether words such as *irrjar* 'three' can refer to groups that contain more than three items, and the answer was consistently and categorically negative.

23. Note that *kujarra* 'two' itself is most likely a borrowing from neighboring Pama-Nyungan languages.

24. This point appears to apply to small numeral systems in other areas of the world, too. Small-limit systems in Amazonia appear to be similarly variable, with a tendency towards nonvague readings. See Epps et al. (2012) for further discussion.

25. This observation is the authors', based on data from the Austronesian Basic Vocabulary Database (http://language.psy.auckland.ac.nz/austronesian/).

26. The word *kunytyi* is also reconstructible in the meaning 'tail', but the semantic association between 'tail' and 'one' is implausible.

27. However, the Worrorran system is somewhat similar to the use of the numeral 'two' as for dual marking; this is found in languages in the Marrngu subgroup of Pama-Nyungan, as well as sporadically elsewhere. None of the other languages in Australia have a numeral system that so clearly relates grammatical number to enumeration.

28. In a recent survey of reconstructions for Pama-Nyungan, ninety out of 195 Pama-Nyungan languages from all subgroups had a reflex of *kutyarra*.

29. Subgroups in which a reflex of *pula is found as a numeral include Maric and Bandjalangic.

30. For example, the word for 'cheese' in Romance languages (French *fromage*, Italian *formaggio*, and so on), comes from the Latin for 'formed'; it was originally the second member of a phrase meaning 'formed cheese' or 'molded cheese'.

31. See also the further discussion contributed by Luise A. Hercus and Isobel M. White as n. 76 of Schebeck's chapter (Schebeck 1973:41–43).

32. The *r* in Arabana *panturu* is a tap.

References

Alpher, Barry
    1991   Yir-Yorront Lexicon: Sketch and Dictionary of an Australian Language. Berlin: Mouton de Gruyter.
Ash, Anna, John Giacon, and Amanda Lissarrague
    2003   Gamilaraay, Yuwaalaraay and Yuwaalayaay Dictionary. Alice Springs, Australia: IAD Press.
Austin, Peter
    1981   A Grammar of Diyari, South Australia. Cambridge: Cambridge University Press.
Beekes, R. S. P.
    1995   Comparative Indo-European Linguistics: An Introduction. Amsterdam: John Benjamins.
Black, Paul
    n.d.   Koko-Berrin. MS. Charles Darwin University.
Blake, Barry J.
    1981   Australian Aboriginal Languages. Sydney: Angus and Robertson.
    1998   Wathawurrung and the Colac Language of Southern Victoria. Pacific Linguistics, Series C, No. 147. Canberra: Research School of Pacific and Asian Studies, Australian National University.
Bowern, Claire
    2012a   A Grammar of Bardi. Berlin: Mouton de Gruyter.
    2012b   The Pama-Nyungan Comparative Lexical Database. MS, Yale University.
Bowern, Claire, and Quentin Atkinson
    2012   Computational Phylogenetics and the Internal Structure of Pama-Nyungan. Language 88:817–45.
Clendon, Mark
    2000   Topics in Worrorra Grammar. Ph.D. diss., University of Adelaide.
Comrie, Bernard
    2005   Numeral Bases. In The World Atlas of Language Structures, edited by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 530–33. Oxford: Oxford University Press.
Corbett, Greville G.
    2000   Number. Cambridge: Cambridge University Press.
Croft, William
    2007   Intonation Units and Grammatical Structure in Wardaman and in Cross-Linguistic Perspective. Australian Journal of Linguistics 27:1–39.
Dawson, James
    1881   Australian Aborigines: The Languages and Customs of Several Tribes of Aborigines in the Western District of Victoria, Australia. Melbourne: G. Robertson.

Dixon, R. M. W.
  1972   The Dyirbal Language of North Queensland. Cambridge: Cambridge University Press.
  1980   The Languages of Australia. Cambridge: Cambridge University Press.
  1991   Words of Our Country: Stories, Place Names, and Vocabulary in Yidiny, the Aboriginal Language of the Cairns-Yarrabah Region. St. Lucia: University of Queensland Press.
  2002   Australian Languages: Their Nature and Development. Cambridge: Cambridge University Press.
Epps, Patience
  2006   Growing a Numeral System: The Historical Development of Numerals in an Amazonian Language Family. Diachronica 23:259–88.
Epps, Patience, Claire Bowern, Cynthia A. Hanson, Jane H. Hill, and Jason Zentz
  2012   On Numeral Complexity in Hunter-Gatherer Languages. Linguistic Typology 16:41–110.
Evans, Nicholas
  1992   Macassan Loanwords in Top End Languages. Australian Journal of Linguistics 12:45–91.
  2003   Context, Culture, and Structuration in the Languages of Australia. Annual Review of Anthropology 32:13–40.
  2009   Two Plus One Makes Thirteen: Senary Numerals in the Morehead-Maro Region. Linguistic Typology 13:321–35.
Evans, Nicholas, and Stephen C. Levinson
  2009   The Myth of Language Universals: Language Diversity and Its Importance for Cognitive Science. Behavioral and Brain Sciences 32:429–48.
Fortson, Benjamin W.
  2004   Indo-European Language and Culture: An Introduction. Malden, Mass.: Blackwell.
Furby, Edward S., and Christine E. Furby
  1977   A Preliminary Analysis of Garawa Phrases and Clauses. Pacific Linguistics, Series B, No. 42. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
Greenberg, Joseph H.
  1978   Generalizations about Numeral Systems. In Universals of Human Language. Vol. 3: Word Structure, edited by Joseph H. Greenberg, Charles A. Ferguson, and Edith Moravcsik, 249–95. Stanford, Calif.: Stanford University Press.
Hale, Kenneth
  1975   Gaps in Grammar and Culture. In Linguistics and Anthropology: In Honor of C. F. Voegelin, edited by M. Dale Kinkade, Kenneth L. Hale, and Oswald Werner, 295–315. Lisse: Peter de Ridder Press.
  1988   Ngarluma Wordlist. Aboriginal Studies Electronic Data Archive (ASEDA) 0037.
Hammarström, Harald
  2008   Numerals in the World's Languages: An Update on Status and Interpretation. http://www.lingvistkredsen.dk/LK_moeder/Oplaeg_Hammerstroem_08.
  2010   Rarities in Numeral Systems. In Rethinking Universals: How Rarities Affect Linguistic Theory, edited by Jan Wohlgemuth and Michael Cysouw, 11–61. Berlin: Mouton de Gruyter.

Hanke, Thomas
  2005   Bildungsweisen von Numeralia: Eine typologische Untersuchung. Berliner
            Beiträge zur Linguistik 3. Berlin: Weissensee Verlag.
  2010   Additional Rarities in the Typology of Numerals. *In* Rethinking Universals:
            How Rarities Affect Linguistic Theory, edited by Jan Wohlgemuth and
            Michael Cysouw, 61–90. Berlin: Mouton de Gruyter.
Harris, John
  1982   Facts and Fallacies of Aboriginal Number Systems. *In* Language and
            Culture, Work Papers of SIL-AAB, Series B, Vol. 8, edited by S. Hargrave,
            153–76. Darwin: Summer Institute of Linguistics, Australian Aborigines
            Branch.
  1987   Australian Aboriginal and Islander Mathematics. Australian Aboriginal
            Studies 2:29–37.
Hercus, Luise A.
  1986   Victorian Languages: A Late Survey. Pacific Linguistics, Series B, No. 77.
            Canberra: Department of Linguistics, Research School of Pacific Studies,
            Australian National University.
  1994   A Grammar of the Arabana-Wangkangurru Language, Lake Eyre Basin,
            South Australia. Pacific Linguistics, Series C, No. 128. Canberra: Depart-
            ment of Linguistics, Research School of Pacific Studies, Australian National
            University.
Hosking, Dianne, and Sally McNicol
  1993   Wiradjuri. Canberra: [privately printed].
Howitt, Alfred William
  1904   The Native Tribes of South-East Australia. New York: Macmillan.
Laughren, Mary, and Robert Hoogenraad
  1996   A Learner's Guide to Warlpiri: Tape Course for Beginners. Alice Springs,
            Australia: IAD Press. (Based on a tape course prepared by Kenneth Hale
            and Robin Japanangka Granites.)
Lynch, John
  1998   Pacific Languages: An Introduction. Honolulu: University of Hawai'i Press.
McConvell, Patrick, and Mary Laughren
  2004   The Ngumpin-Yapa Subgroup. *In* Australian Languages: Classification and
            the Comparative Method, edited by Claire Bowern and Harold Koch,
            151–77. Amsterdam: John Benjamins.
McEntee, John C., and Pearl McKenzie
  1992   Adna-mat-na English Dictionary [Adna-mat-na/English]. Adelaide:
            [privately printed].
McGregor, William
  1990   A Functional Grammar of Gooniyandi. Amsterdam: John Benjamins.
  2004   The Languages of the Kimberley, Western Australia. London: Routledge-
            Curzon.
McRoberts, Robert
  1990   Counting at Pularumpi: A Survey of a Traditional Mathematics and Its
            Implications for Modern Learning. Aboriginal Child at School 18:19–43.
Meeussen, A. E.
  1967   Bantu Grammatical Reconstructions. Africana Linguistica 3:79–128.
Mitchell, T. L.
  1848   Journal of an Expedition into the Interior of Tropical Australia, in Search of
            a Route from Sydney to the Gulf of Carpentaria. London: Longman, Brown,
            Green and Longmans.

Nekes, Hermann, and Ernest A. Worms
    2006   Australian Languages. Edited by William B. McGregor. Berlin: Mouton de
            Gruyter.
Oates, Lynette F.
    1988   The Muruwari Language. Pacific Linguistics, Series C, No. 108. Canberra:
            Department of Linguistics, Research School of Pacific Studies, Australian
            National University.
O'Grady, Geoffrey N., C. F. Voegelin, and F. M. Voegelin
    1966   Languages of the World: Indo-Pacific Fascicle 6. Anthropological Linguis-
            tics 8:1–199.
Osborne, C. R.
    1974   The Tiwi Language. Canberra: Australian Institute of Aboriginal Studies.
Peile, Anthony Rex
    1993   A Basic Kukatja to English Dictionary, edited by Hilaire Valiquette. Abori-
            ginal Studies Electronic Data Archive (ASEDA) 0504. Wirramanu (Balgo),
            Australia: Lurnpa Catholic School.
Ray, Sidney H., and Alfred C. Haddon
    1891   A Study of the Languages of Torres Straits, with Vocabularies and Gram-
            matical Notes (Part I). Proceedings of the Royal Irish Academy, 3d series,
            2:463–616.
Rhodes, Richard, and David J. Costa
    2003   The History of Algonquian Number Words. In Essays in Algonquian,
            Catawban and Siouan Linguistics in Memory of Frank T. Siebert, Jr.,
            edited by Blair A. Rudes, David J. Costa, and Frank T. Siebert, 181–215.
            Winnipeg: Algonquian and Iroquoian Linguistics.
Richards, Eirlys G., and Joyce Hudson
    1990   Walmajarri-English Dictionary with English Finder List. Darwin, Aus-
            tralia: Summer Institute of Linguistics.
Ridley, William
    1875   Kamilaroi and Other Australian Languages. 2d ed. Sydney: T. Richards,
            Government Printer.
Robertson, Carol
    1984   Wangkumara Grammar and Vocabulary. Sydney: Department of Technical
            and Further Education, Aboriginal Education Unit.
Saulwick, Adam
    2003   A First Dictionary of Rembarrnga. Maningrida, Australia: Bawinanga
            Aboriginal Corporation.
Schebeck, B.
    1973   The Adnjamathanha Personal Pronoun and the "Wailpi Kinship System."
            In Papers in Australian Linguistics 6, 1–45. Pacific Linguistics, Series A,
            No. 36. Canberra: Department of Linguistics, Research School of Pacific
            Studies, Australian National University.
Simpson, Jane, and Luise Hercus
    2004   Thura-Yura as a Subgroup. In Australian Languages: Classification and
            the Comparative Method, edited by Claire Bowern and Harold Koch, 179–
            206. Amsterdam: John Benjamins.
Simpson, Jane Helen
    2002   A Learner's Guide to Warumungu: Mirlamirlajinjjiki Warumunguku
            apparrka. Alice Springs, Australia: IAD Press.
Stokes, Bronwyn
    1982   A Description of Nyigina: A Language of the West Kimberley, Western
            Australia. Ph.D. diss., Australian National University, Canberra.

Tsunoda, Tasaku
    1972   Warungu Dictionary. MS.
    1981   The Djaru Language of Kimberley, Western Australia. Pacific Linguistics, Series B, No. 78. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
    2011   A Grammar of Warrongo. Berlin: Mouton de Gruyter.

Tylor, Edward B.
    1871   Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom. London: John Murray.

Waters, Bruce E.
    1989   Djinang and Djinba: A Grammatical and Historical Perspective. Pacific Linguistics, Series C, No. 114. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

Yallop, Colin
    1982   Australian Aboriginal Languages. London: A. Deutsch.

Zerbian, Sabine, and Manfred Krifka
    2008   Quantification across Bantu Languages. *In* Quantification: A Cross-Linguistic Perspective, edited by Lisa Matthewson, 383–414. North-Holland Linguistic Series, Linguistic Variations 64. Bingley, U.K.: Emerald.

Zorc, R. David
    1986   Yolngu-Matha Dictionary. Batchelor: School of Australian Linguistics, Darwin Institute of Technology.