Did you have a choccie bickie this arvo? A quantitative look at Australian hypocoristics.

Evan Kidd

*La Trobe University & The University of Manchester*

Nenagh Kemp & Sara Quinn

*University of Tasmania*

Contact details:

Evan Kidd
School of Psychological Sciences
La Trobe University
Bundoora 3086
Victoria AUSTRALIA
Ph: +61 3 9479 2420
Fax: +61 3 9479 1956
Email: e.kidd@latrobe.edu.au
Abstract

The current paper considers the use and representation of Australian Hypocorositics (e.g., *choccie* → *chocolate*, *arvo* → *afternoon*). One-hundred-and-fifteen adult speakers of Australian English aged 17 – 84-years generated as many tokens of hypocoristics as they could in ten minutes. The resulting corpus was analysed along a number of dimensions in an attempt to identify (i) general age- and gender-related trends in hypocoristic knowledge and use, and (ii) linguistic properties of each hypocoristic class. Following Bybee’s (1985, 1995) lexical network approach, we conclude that Australian hypocoristics are the product of the same linguistic processes that capture other inflectional morphological processes.
Introduction

Much has been made of the peculiarities of Australian English (AusE). It has often been claimed that the local dialect can be rather difficult for non-AusE speakers to understand (e.g., McAndrew, 1992). Such difficulty cannot solely be attributed to differences in accent, since the language contains literally thousands of lexical items that are alternative to the standard English forms. Indeed, there is a veritable cottage industry of dictionaries of Australian slang, beginning with Professor Afferbeck Lauder’s (1965) tongue-in-cheek *Let Stalk Strine*, and culminating in publications such as the *Dinkum Dictionary* (Butler, 2009) and the *Macquarie Slang Dictionary* (Lambert, 2004), which are produced as much for humorous as documentation purposes.

The current paper focuses on Australian *hypocoristics*, such as *choccie*, *arvo*, and *brekkie*. These are colloquial forms that have the same denotation and often share some of the same form as the standard word that they denote, but differ from the standard form in their formality and sometimes in their connotation. Most are characterised by the fact that they clip the base form and add a morpheme, usually –*o*, –*ie/y*, –*a/er*, or the more recent –*s*, as shown in (1) – (4).

1. service station → servo.
2. cigarette → coggie.
3. sandwich → sanga.
4. mobile phone → mobes.

In spoken AusE many of these terms are used with higher frequency than their standard forms, and yet their use is often derided for being uneducated and unrefined.

---

1 Professor Afferbeck Lauder was the nom de plume of Alistair Morrison, an artist and humorist. When pronounced in a heavily accented Australian accent his pseudonym sounds like ‘alphabetical order’.
2 Others such as McAndrew (1992) and Wierzbicka (1986) have called these forms ‘diminutives’; however, we follow Simpson (2001, 2004) in calling them ‘hypocoristics’.
A quantitative look at Australian hypocoristics

(Kidd, Kemp, Kashima, & Kashima, in preparation; McAndrew, 1992; for a brief historical perspective see Sussex, 2004). In this paper we argue that, far from being a quirk of the dialect of a restricted category of speakers (i.e., the uneducated and unrefined), this class of words are known and used by a majority of the Australian speaking community, and are in many ways subject to same types of morphological processing as standard words.

Past research on hypocoristics

Moore (2008) notes that the first recorded instances of hypocoristics were in the 19th century, with exemplars of both the –o and –ie morpheme. The earliest recorded use of the –o morpheme was in spell-o, ‘a call signalling a break from work’ (later replaced by smoko), and the earliest recorded of the –ie morpheme (in AusE) was schoolie, which was used to refer to a school teacher. The 20th century saw an explosion of the use of these forms: Simpson (2004) reports on an analysis of 1740 different forms collected over a number of decades from a variety of sources (e.g., directly from speakers, other publications, media, and general observation). More comprehensively still, Sussex (2004) reports having collected some 4,300 forms, estimating that hypocoristics make up some 4% of the types of Australian lexis. Such reports suggest that the use of the hypocoristic morphemes became productive over the course of the last century. Below we review some of the past attempts to categorise these forms. In our discussion we exclude proper and place names, which have been considered in detail elsewhere (Dabke, 1976; Mühlhäusler, 1983; Simpson, 2001; Taylor, 1992; Wierzbicka, 1986).

Dabke (1976) provided one of the first comprehensive studies of the morphology of AusE, in which she considered hypocoristics. In particular, she

---

3 Although Sussex (2004) reports to have documented 4,300 forms, the actual number of head words that can be used to form hypocoristics will be less than this number, since some heads take multiple hypocoristic morphemes (e.g., afternoon → arvo, aftie, arv, afto).
A quantitative look at Australian hypocoristics

described the use of the –ie and –o morphemes, gleaned from observation and interviews with 11 native speakers from various backgrounds. Although her analysis was purely descriptive, it provides us with a starting point from which to work. Structurally, Dabke distinguished between base forms that are clipped before a suffix is added (e.g., commie → communist; journo → journalist), and those that add the target morpheme to the base word, which is typically one syllable in length (e.g., kiddie → kid). She argued that in the former case suffixation does not alter the meaning of the base form, whereas in the latter case it often does. She also noted the apparent flexibility of the morphemes which, although mostly used to modify nouns, can be derived from other parts of speech, including verbs and adjectives. Finally, Dabke attempted to identify the major semantic fields to which the hypocoristics she observed belonged. She lists specific occupations (e.g., postie → postman, garbo → garbage collector), terms for sportspeople (e.g., surfie → surfboard rider, yachty → person who sails yachts), an animal’s condition (e.g., gummy → without teeth, dopey → tired, slow), and, rather broadly, objects and actions.

Wierzbicka (1986) wrote briefly about hypocorositics, which she called ‘depreciatives’, and attempted to formulate precise semantic meaning for both the –ie and –o suffixes using Natural Semantic Metalanguage (NSM). Specifically, she argued for the following definitions of each (5 & 6).

(5) –ie (e.g., prezzie, mozzie)

I don’t want to think much about X
I think that you think of X like I do
Speaking about it with you I feel something special.
I don’t think of it like people think of things they don’t know about
I think you think of it like I do
Speaking of it I don’t want to use long words
Therefore, Wierzbicka suggested that these two morphemes can be
distinguished by the fact that the –ie morpheme is used in cases where the object or
action in question is downgraded in importance, and where a speaker selects an
informal register to establish common ground. In contrast, she suggests that that the –
o morpheme is used when speakers talk of common and mutually understood topics,
and therefore prefer to shorten their pronunciation.
McAndrew (1992) also discussed the –ie and –o morphemes. Like Wierzbicka
(1986), he attempted to categorise the two into broad semantic classes, but also
discussed subclasses. He referred to words in the –ie class as diminutives or
‘lovables’, and words in the –o class as pejoratives or ‘unlovables’. Let us consider
this categorisation in more detail.
McAndrew (1992) argues that, in general, words belonging to the –ie class
represent “mostly elegant, affectionate, and familiar” concepts (p. 174). Within this
superordinate category he describes five subclasses. Words in the first subclass are
argued to denote ‘expressions of mateship and affection’, which largely includes
terms for people and vocations; for instance, wharfie (waterside worker), tradie
(tradesperson), sparkie (electrician), and yuppie (young urban professional). Words in
the second subclass denote ‘terms of familiarity and informality’, which include
largely inanimate objects or animals common to Australian life, such as cozzie
(swimming costume), *barbie* (barbecue), *maggie* (magpie), and *pokies* (poker machines). The third subclass McAndrew calls ‘terms of rebuke’, which include terms like *pollie* (politician), *blowie* (blowfly), and *bikie* (motorcyclist). The fourth subclass are ‘mirage suffixes’, which, although they conform to the general structure of this type of hypocoristic (in general, disyllabic and ending in –*ie*), are not hypocoristics because they do not undergo any morphological change. Examples include *humpy* (a small make-shift shelter) and *gladdie* (gladiolus). The final class are adjectival nouns such as *smoothie*, *quickie*, and *toughie*, which tend to act as shorthand for a pronominal adjectival construction (e.g., ‘that’s a *tough one*’).

In contrast to the –*ie* suffix, McAndrew (1992) argues that the ‘unlovable’ terms formed using the –*o* suffix are ‘coarse, vigorous, excessive, and used more by males’ (p. 174). He argues that words in this superordinate category fall into five subclasses. The first subclass is ‘terms of contempt and ridicule’, which includes terms like *reffo* (refugee), *abo* (aboriginal), *dago* (Southern European), and *homo* (homosexual). The second subclass is ‘terms of laziness and carelessness’. This category includes terms such as *garbo* (garbage collector) and *demo* (demonstration). The third category is ‘terms of excess’, which includes terms such as *aggro* (aggressive), *berko* (to go beserk), and *wino* (alcoholic). The fourth is, once again, ‘mirage suffixes’ such as *smoko* (work break), *demo* (demonstration), and *physio* (physiotherapist). The final rather small category McAndrew calls the ‘Tall Poppy Syndrome’, in which, in keeping with Australian tradition, he argues that the –*o* suffix is used to criticise public figures and movements (e.g., *Cathos* for Catholics).

Although McAndrew’s (1992) analysis is wide-reaching, in reality these categories have fairly fuzzy boundaries that seem to resist neat categorisation. For instance, *yuppie* is more often used in a derogatory sense than to express mateship and
affection. Similarly, the argument that terms such as *garbo* and *demo* are used because speakers of AusE are lazy and careless and like to shorten words should not just apply to one category, but to the whole class of hypocoristics in general, thereby leaving the members of this class without a home. Indeed, McAnd...
hypocoristics. Clearly speakers of AusE treat these as the same as hypocoristics that require morphological composition. What the templatic approach does not predict, however, is morpheme selection. That is, what leads to the selection of the \(-a\) morpheme over the \(-ie\) morpheme? We attempt to address this question in the current study.

[insert Table 1 about here]

The current study.

The current research had several aims. Firstly, we aimed to collect a corpus of hypocoristics from a large cross-sectional sample of speakers of AusE. Previous studies of hypocoristics have used opportunistic sampling, collecting tokens when they arose over long periods; by asking a large number of people to produce hypocoristics we hoped to add to the literature by documenting current trends in hypocorotic use in AusE across several generations of speakers. Secondly, using our corpus we aimed to (i) evaluate the past analyses of Australian hypocoristics, (ii) analyse the corpus for age- and gender-related trends, and (iii) analyse any discernible linguistic features of the hypocoristic types that we observed.

Method

Participants

The participants were 115 individuals aged from 17 to 84 years (\(M = 41.2\) years, \(SD = 21.1\) years), living in South-Eastern Australia (Hobart & Melbourne). Most participants (94.8\%) had lived in Australia all their lives, and the remaining had lived in Australia, on average, for half of their lives (\(M = .50\); \(SD = .35\)). For the purposes of analyses the sample was divided into three ‘generational’ groups: (i) 17 – 39 year-olds \((N = 61, 15\) males, 46 females), who had a mean age of 24.0 years \((SD =\)
A quantitative look at Australian hypocoristics

6.5 years), (ii) 40 – 59-year-olds ($N = 27$, 9 males, 18 females), who had a mean age of 49.4 years ($SD = 6.3$ years), and (iii) 60 – 84 year-olds ($N = 27$, 8 males, 19 females), who had a mean age of 72.1 years ($SD = 6.7$ years). We refer to the youngest group as ‘Young’, the middle age group as ‘Middle-aged’, and the eldest age group as ‘Older’. The participants were either undergraduate Psychology students ($n = 41$) who participated for course credit, or friends and work colleagues of the authors. Since the sample was recruited using an opportunistic method, there were more female than male respondents (82: 33). Thus all of the analyses involving gender were performed on percentage rather than numerical data. With the exception of the Young group, who almost exclusively had completed or were undertaking university education, our sample was drawn from a broad cross-section of Australian society, with a range of educational and socio-economic backgrounds.

*Materials and Procedure*

Participants were given several examples of the kinds of abbreviations that are often used in Australian English (e.g., *arvo* for *afternoon*; *choccie* for *chocolate*). They were then asked to write down as many such abbreviations as they could think of in ten minutes on a numbered sheet of paper. Participants completed this exercise in their own time and returned their response sheets to the experimenters. The data were collected between May 2009 and April 2010.

**Results**

The participants’ responses were coded according to Simpson’s (2004) categories, as outlined in Table 1. There were, however, a few exceptions. As we were only concerned with common nouns, we did not include Simpson’s *The* category (e.g., *The Don*, for Donald Bradman), since it is almost exclusively used for proper
nouns. Secondly, we did not include acronyms, since we were more interested in the morphological processes involved in hypocoristic formation.

**Descriptives**

The Young age group provided an average of 25.9 \((SD = 16.5)\) hypocoristic tokens. The Middle-aged group provided an average of 26.1 \((SD = 16.4)\), and the Older group provided an average of 17.9 \((SD = 10)\). Females \((M = 24.2, SD = 16.9)\) provided more tokens on average than did males \((M = 20.9, SD = 10.7)\). This trend was true for both the Young (Females: \(M = 27.9, SD = 17.7\); Males: \(M = 19.7, SD = 9.9\)) and Middle-Aged groups (Females: \(M = 28.2, SD = 19.3\); Males: \(M = 21.9, SD = 7.0\)), whereas the opposite was the case for the Older group (Females: \(M = 16.4, SD = 6.1\); Males: \(M = 21.0, SD = 15.3\)). However, a series of non-parametric Kruskal-Wallis and Mann-Whitney \(U\)-tests showed that these gender differences were not statistically significant for the entire sample, and nor were they for individual age groups. There were, however, age differences in the data \((\chi^2 = 7.15, df = 2, p = .028)\).

Post-hoc Mann-Whitney \(U\)-tests showed that, whereas the Young and Middle-aged group did not differ from each other \((U = 807.5, p = .89)\), both of these groups provided significantly more hypocoristic tokens than did the Older group (Young > Older: \(U = 586.5, p = .032\), Middle-aged > Older: \(U = 201.0, p = .005\)).

Table 2 shows the distribution of hypocoristic types for each age group.

[insert Table 2 about here]

Table 2 shows that the distribution of hypocoristic types differs slightly across the three generations of speakers. Three trends are immediately obvious. Firstly, there is an increase in the number of \(-ie\) and \(-o\) forms with age, suggesting that these two
morphemes may be used less in younger generations. In contrast, the use of one-syllable and two-syllable types decreased with age, suggesting that these forms may be gaining ground on the two more traditional hypocoristic forms. Finally, we point out that although the Young age group provided many more types than did the other two generational groups, it does not necessarily mean that there are more forms in use among young people. In fact, when the number of participants in each group is controlled, the Young age group provided an average of 6.1 individual types, whereas the Middle-aged and the Older groups provided an average of 7.2 and 6.4, respectively.

Next we present the type-token ratios for each category by age group. Type-token ratio is a measure of lexical diversity within a particular category. It is simply calculated by dividing the number of types by the number of tokens for any one category. Since the number of types will always be equal to or less than the number of tokens, the type-token ratio will always be a number between 0 and 1. The higher the type-token ratio, the greater the lexical diversity in that category. Table 3 lists the type-token ratios for each hypocoristic category by age group.

[insert Table 3 about here]

Table 3 shows that hypocoristic use within each category is similar across the three age groups; all groups produced examples of each category at around the same rate and with the same degree of within category diversity. At first blush this suggests fairly stable use across the three generations. The one exception appears to be the -a form, which is used more often and with more diversity by the Young age group. We now consider each category in more detail.
This is the largest class of hypocoristics in terms of both type and token frequency, consistent with other findings by Simpson (2004). The 10 most frequent forms were *footy → football* (43), *brekkie → breakfast* (34), *telly → television* (33), *bickie → biscuit* (31), *veggie → vegetable* (21), *barbie → barbeque* (20), *ciggie → cigarette* (15), *sunnies → sunglasses* (15), *mozzie → mosquito* (13), and *undies → underwear* (13). All of these forms were very common across the three age groups. The group is made up almost exclusively of nouns, although there were three forms that can be used as adjectives (*guvvie → government, exi → expensive, iffy → suspect*). Semantically, this group is what McAndrew (1992) called the “loveables”, due to their tendency to associated with friendship and familiarity. Wierzbicka’s (1986) analysed this category as similarly positive. To be sure, this category includes four distinctly affectionate terms of endearment (*lovey, matie, sweetie, missie*). However, this morpheme is most often used in two contexts: for artifacts (85/184 = 46.2%) (e.g., *bickie → biscuit, boardies → board shorts, pressie → present, tinnie → can of beer*) and to describe people (56/184 = 30.1%), mainly occupations (e.g., *carnie → carnival worker, chippie → carpenter, hostie → hostess, sparkie → electrician*). Although mostly affectionate, some terms have clear negative connotations (e.g., *druggie → drug addict, parkie → aboriginal person who lives in an urban park*). However, our data mostly concur with the arguments made by McAndrew and Wierzbicka. Wierzbicka’s analysis should almost certainly be taken to refer to the pragmatics of hypocoristic use in addition to semantics, in which case forms that have clearly negative connotations still function to afford a level of familiarity and shared perspective between speakers.
The other group with a substantial number of members were terms used to describe events \((15/184 = 8.2\%)\) (e.g., \(\text{Chrissie} \rightarrow \text{Christmas}, \text{footie} \rightarrow \text{football}, \text{pissie} \rightarrow \text{‘a piss-up’}\) or a party where people drink a lot of alcohol, \(\text{walkies} \rightarrow \text{walking the dog}\)).

\(-o\) morpheme

The \(-o\) morpheme is what McAndrew (1992) the ‘unlovables’, so called because they are purported to be “coarse, vigorous, excessive, and used more by males” (pp. 174). The 10 most frequent forms were \(\text{servo} \rightarrow \text{service station}\) (18), \(\text{garbo} \rightarrow \text{garbage collector}\) (17), \(\text{arvo} \rightarrow \text{afternoon}\) (13), \(\text{ambo} \rightarrow \text{ambulance worker}\) (11), \(\text{smoko} \rightarrow \text{break from work}\) (10), \(\text{bottle} \rightarrow \text{liquor store}\) (or ‘bottle shop’) (9), \(\text{derro} \rightarrow \text{derelict person}\) (8), \(\text{muso} \rightarrow \text{musician}\) (8), and \(\text{milko} \rightarrow \text{milkman}\) (6). Rather than being purely ‘unloveable’ terms, the majority of these high frequency forms encode occupations or describe classes of people. This trend is mirrored in the category in general, where \(25/56\) (44.6\%) of the \(-o\) suffix hypocoristics describe a category of person. Just over half of these (14) have some negative connotations or uses (e.g., \(\text{alco} \rightarrow \text{alcoholic}, \text{lezzo} \rightarrow \text{lesbian}, \text{derro} \rightarrow \text{derelict}, \text{povo} \rightarrow \text{poverty-stricken}\)), but the remainder do not (e.g., \(\text{ambo} \rightarrow \text{ambulance worker}, \text{journo} \rightarrow \text{journalist}, \text{deso} \rightarrow \text{designated driver}, \text{muso} \rightarrow \text{musician}\)). The other two major category within this class are terms for artifacts (14/56 = 25\%, e.g., \(\text{doco} \rightarrow \text{documentary}, \text{flanno} \rightarrow \text{flannelette shirt}, \text{metho} \rightarrow \text{methylated spirits}, \text{rego} \rightarrow \text{registration}\)) and states (6/56 = 10.7\%, e.g., \(\text{aggro} \rightarrow \text{aggressive}, \text{berko} \rightarrow \text{beserk/angry}, \text{troppo} \rightarrow \text{crazy}\)). This latter category most often describe negative attributes.

McAndrew (1992) observed that words that take the \(-i\) morpheme and ones that take the \(-o\) morpheme to form hypocoristics have different etymological roots. In
particular, those that take the –i morpheme have Germanic roots, whereas those that take the –o morpheme have Latinate roots. Indeed, as McAndrew pointed out, these morphemes have roots as diminutives in German and French. The average Australian, of course, does not speak enough of either language to argue for a direct effect of either language on hypocoristic formation or use. McAndrew (1992) suggested that instead it must be the word length of the base word that contributes to morpheme selection, since Latinate words are generally longer than Germanic words. We tested this prediction by statistically comparing the number of syllables of each base word between these two categories. For this analysis only words that entered into putatively ‘regular’ morphological processes were included (e.g., *alcoholic* → *alchie/alcho*, *derelict* → *derro*), since irregular forms such as *carpenter* → *chippie* are not predictable from their base forms. McAndrew’s prediction was borne out: the base forms of words that take the –o morpheme (M = 3.1 syllables, SD = 1.4) contained significantly more syllables than did words that take the –ie morpheme (M = 2.5, SD = .96), U = 3503.0, p = .002. In contrast, there was no difference when the number of syllables of the hypocoristic forms of the two groups was compared (U = 4649.5, p = .26). This is because, consistent with observations made by Simpson (2004), both categories tended to be accommodated within a 2-syllable template (-o morpheme: M = 2.04, SD = .19, -ie morpheme: M = 2.01, SD = .11).

The selection of the two morphemes can also be distinguished by the frequency with which they encode the two main semantic categories: people and artifacts. A chi-square analysis showed this difference to be significant ($\chi^2 = 7.53$, $p < .01$). That is, words that take the –i morpheme are more likely to encode an artifact, whereas words that take the –o morpheme are more likely to encode a person or attribute.
Finally, we tested whether McAndrew’s (1992) prediction that males use more ‘unloveables’ than do females. Overall, this did not seem to be the case (Females: \(M = 1.7\), Males: \(M = 1.97\)). However, use of this form interacted with age. Whereas Young females (\(M = 1.41\)) listed more than did Young males (\(M = 1.07\)), Middled-aged males (\(M = 3.0\)) provided twice as many on average than did Middle-aged females (\(M = 1.5\)). Older males (\(M = 2.75\)) provided slightly more than did Older females (\(M = 2.5\)).

One syllable

The one-syllable hypocoristics are, by definition, abbreviations of the base forms, corresponding to Simpson’s (2004) second template category. The most frequent in our corpus were \(psych \rightarrow psychology\) (10), \(comp \rightarrow competition/computer\) (9), \(sis \rightarrow sister\) (9), \(bro \rightarrow brother\) (8), \(mag \rightarrow magazine\) (8), \(ref \rightarrow referee\) (8), \(doc \rightarrow doctor\) (6), \(mic \rightarrow microphone\) (6), and \(ump \rightarrow umpire\) (6). As shown by the type-token ratio analysis, this category had many low frequency members. Indeed, our highest frequency example \(psych\) is almost certainly due to the fact that many of our participants were psychology students or worked in psychology departments. This category also seems to be growing in use: whereas the Older group only provided an average of 1 token each, the Middle-aged and Young group provided an average of 2.2 and 3.1, respectively.

Two syllable

Like the one-syllable category, the two-syllable category had quite a few mostly low frequency members. The most frequent items were \(uni \rightarrow university\) (18), \(info \rightarrow information\) (6), and \(demo \rightarrow demonstration\) (4). Again, there were age effects in the data: participants in the Young (\(M = 1.02\)) and Middle-aged (\(M = 0.96\)) groups
provided on average an equivalent number, whereas the Older group provided fewer on average \((M = 0.6)\).

\textit{-a morpheme}

The \textit{-a} morpheme category is marked by the large number of irregular forms inhabiting this category. Twenty-seven out of 35 tokens (\(= 77\%\)) were either suppletive (e.g., \textit{banger} \(\rightarrow\) \textit{sausage}, \textit{beefr} \(\rightarrow\) \textit{sexually promiscuous female}, \textit{diller} \(\rightarrow\) \textit{franger} \(\rightarrow\) \textit{condom}) or require a stem change (e.g., \textit{mozzer} \(\rightarrow\) \textit{mosquito}, \textit{nowukkers} \(\rightarrow\) \textit{no worries}, \textit{sanga} \(\rightarrow\) \textit{sandwich}, \textit{stezza} \(\rightarrow\) \textit{stereo}). This category only had two high frequency members: \textit{cuppa} \(\rightarrow\) \textit{cup of tea} (23) and \textit{sanga} (15). The use of this category was evenly distributed across the three age groups, with participants in the Young group providing an average of 0.85 tokens, the Middle-aged participants 0.56, and the Older group 0.7.

\textit{-as category}

This very small category, containing only 3 members: \textit{champers} \(\rightarrow\) \textit{champagne} (3), \textit{preggers} \(\rightarrow\) \textit{pregnant} (1), and \textit{snippers} \(\rightarrow\) \textit{scissors} (1). Although not included in our analyses, this category seems to be almost exclusively reserved for proper nouns (e.g., \textit{Sydders} \(\rightarrow\) Sydney, \textit{Honkers} \(\rightarrow\) Hong Kong, \textit{Sainters} \(\rightarrow\) St Kilda Football Club).

\textit{Syllable + -s category}

The syllable + \textit{-s} category is marked by a large number of monosyllabic forms: 21/28 (77\%) in total. The most frequent members in the category were: \textit{maybs} \(\rightarrow\) \textit{maybe} (4), \textit{mobes} \(\rightarrow\) \textit{mobile phone} (4), \textit{peeps} \(\rightarrow\) \textit{people} (4), \textit{fuggers} \(\rightarrow\) \textit{very ugly} (4), \textit{whatevs} \(\rightarrow\) \textit{whatever} (4), \textit{defs} \(\rightarrow\) \textit{definitely} (3), and \textit{probs} \(\rightarrow\) \textit{probably} (3). These forms were more common in the Young \((M = 0.54)\) and Middle-aged groups \((M = 0.37)\) than in the Older group \((M = 0.07)\).
This is another small category that was more common in the Young ($M = 0.56$) and Middle-aged ($M = 0.56$) groups than in the Older group ($M = 0.15$). The most frequent forms were $roo \rightarrow kangaroo$ (13), $nana \rightarrow banana$ (6), and $puter \rightarrow computer$. The hypocoristic forms were almost evenly split between 1 (54%) and 2 syllable (46%) forms.

Discussion

In the current paper we have reported on the first quantitative study of Australian hypocoristics conducted with a randomly selected sample of speakers of AusE. Our aim was to collect a representative sample of hypocoristic use from a large cross-sectional sample of speakers, and to use the sample to (i) identify usage trends across our defined age groups, (ii) evaluate some of the hypotheses that have been proposed in the literature but have not been tested, and (iii) to identify any other discernible trends in the data. We first discuss our main findings, and then consider the wider linguistic implications of this phenomenon.

The first and most general conclusion we can draw from our data is that people across a broad range of ages and backgrounds know and presumably use hypocoristics. We observed some age-related differences in the use of hypocoristics. In particular, whereas it seems that the $-o$ form is being used less by younger speakers than by older speakers, younger speakers used more one- and two-syllable forms. The reason for this is unclear. It could be that the $-o$ form is slowly falling out of usage. A closer inspection of the data from the Young age group provides support for this assertion. If we divide this group into two to form (i) 17 – 30-year-olds, and (ii) 31 – 40-year-olds, we find a striking difference in their overall usage of the $-o$ morpheme. The 17 – 30-year-olds provided an average 0.76 tokens of this form, whereas the 31 –
40-year-old group provided an average of 3.3 tokens. The opposite trend was observed for the one-syllable forms: the 17 – 30-year-old group provided an average of 3.8 tokens, whereas the 31 – 40-year-old group provided an average of 1.73 tokens. With one exception (def ↔ defo → definitely), there is no overlap in the words that inhabit these two categories, which suggests that the change in the youngest participants may be either a stylistic preference or simply due to the content of their everyday usage. That is to say, the youngest group generally use terms such as diff → difference, hon → honey, and pash → passionate (kiss) more than they use terms such as ambo → ambulance driver, derro → derelict, and prego → pregnant. This is a matter for future sociolinguistic study.

We also provided support for McAndrew’s (1992) prediction that word length affects the selection of the –ie and –o morpheme. Additionally, we also showed that the two categories differ in the extent to which they encode artifacts (-ie > -o) and human-related categories (-o > -ie). This suggests that the –o suffix most commonly acts as an agentive marker. In contrast, the –ie suffix most commonly functions in a manner closest to traditional diminutive insofar as, following Wierzbicka (1986), their use conveys interpersonal warmth while diminutivising the object of interest. Of course, the –ie suffix also has an agentive function (e.g., druggie → drug addict, groupie → band follower, foodie → person with a deep interest in food). These forms most likely take the –ie suffix because their base forms are short, following the general trend for base form length being a good predictor of morpheme type, at least between these two morphemes, which are the two most common hypocoristics that enter into morphological composition. These observations lead us to suggest that the use of different hypocoristic morphemes must be dually determined by morphophonemic and semantic processes. We expand on this point below. Our
argument is that, although Australian hypocoristics appear to constitute a messy linguistic category that defy neat analysis, they exhibit signature properties common to other inflectional systems.

Any attempt to account for the mental representation of Australian hypocoristic forms must deal with the following (messy) facts. Firstly, although there are tendencies, no single morphological process in this class can be said to map onto a unique semantic attribute. Secondly, any explanation must accommodate the fact that some forms have multiple hypocoristic forms (e.g., garbie ↔ garbo → garbage collector, commie ↔ commo → communist, preggers ↔ prego → pregnant), many of which do not differ discernibly in meaning. Bybee’s (1985, 1988, 1995) lexical network model of morphology provides a solid theoretical basis that accommodates these potential problems. The model argues that morphological structure is an emergent property of the organisational structure of the lexicon. Lexical items are stored according to two properties: semantics and phonology. Figure 1 shows a simplified example using the lexical items cat, cats, mats.

[insert Figure 1 about here]

In Figure 1 the thick lines illustrate a strong phonological and semantic overlap between phonemes. The thinner lines denote phonological overlap. We suggest that Australian hypocoristics are naturally explained by Bybee’s model as lexical entries that are connected to both their base forms and other hypocoristics. We argue for this theoretical treatment by pointing out that hypocoristics share one striking similarity with other inflectional processes – a cline of regularity that interacts with semantic transparency. Consider, for instance, English past tense morphology.
Just as we have regular past forms (e.g., *walk* → *walked*), zero-marked forms (e.g., *hit* → *hit*), irregular stem change past forms (e.g., *ring* → *rang*), and suppletive past forms (e.g., *go* → *went*), we have the same processes in hypocoristic formation. Regular processes such as *cigarette* → *ciggie*, *nightdress* → *nightie*, *alcoholic* → *alco*, *registration* → *rego*, *mobile* → *mobes*, *probably* → *probs* are directly predictable from their stems. As we have shown, regular hypocoristics that take either the –*ie* or –*o* morpheme conform to Simpson’s (2004) two-syllable template, and morpheme selection is determined by the length of the base form, in addition to the probable semantic factors described by McAndrew (1992) and Wierzbicka (1986). The syllable +*s* category differs from the previous two insofar as the hypocoristic forms are mostly one syllable in length, and the category itself seems to be expanding. As such, there are likely to be connections between the –*ie* and –*o* forms, but perhaps less so between these forms and the syllable +*s* category. Consider Figure 2.

**Figure 2** shows a hypothesised lexical network for the hypocoristics *connie*, *commie*, *cozzie*, *commo*, *combo*, and *compo*. We see that all six forms share the same onset, and individually share a number of other phonological features, such as geminates or bilabial consonants in their second syllable. The network itself unifies the two hypocoristic morphemes through the different denotations of *communist*, thus illustrating a semantic connection. In Figure 2 we make two links between the two morphemes. The first is between the two hypocoristic forms *commie* and *commo*. This link is a solid line that is intended to indicate (i) semantic overlap insofar as they are both hypocoristic morphemes that in this instance mean something very similar, and
(ii) phonological overlap insofar as both morphemes are vowels. The second link is between the *lil coZZie* and *tho lol* in *commo*. This is a broken line that indicates the same degree of phonological overlap as observed in the previous case, but a lesser degree of semantic overlap. That is, in this case the two morphemes both belong to the superordinate class of hypocoristic morphemes. Note that the link between the two morphemes in *coZZie* and *commo* has been arbitrarily selected; a fully specified schematic would also have links between the two morphemes in *combo* and *compo*.

Irregular forms abound among Australian hypocoristics, most of which are suppletive. Among these suppletive forms there are islands of both phonological and semantic regularity. One particularly notable group includes *chippie* → carpenter, *sparkie* → electrician, *wharfie* → waterside worker, *brickie* → bricklayer, and *navvie* → sailor. These forms share phonological similarity insofar as they all conform to Simpson’s (2004) 2 syllable + -ie template, and are also semantically similar, denoting vocational trades. As such, they denote a product-oriented schema insofar as their phonological similarity is only evident in the hypocoristic and not the base forms. Their hypocoristic form is predictable from semantic associations with their base form, which in the above cases is a semantically notable feature of each profession. Such semantic associations in these ‘irregular’ forms extend beyond blue collar professions; for instance, *digger* → Australian soldier, *ranga* → red-haired person (from orangutan), and *tinker* → inventor. These examples are consistent with current connectionist models of English morphology (Joanisse & Seidenberg, 1999), which learn regular and irregular forms by placing differential importance on encoding either phonological (for regular forms) or semantic information (for

---

5 There is also the very common *cop* → policeman, which has a long history in English. The precise source of this innovation is unclear, although one hypothesis is that it derives from the colloquial verb *to cop* (i.e., to catch/take).
irregular forms). Our suggestion here is that Australian hypocoristics are formed by the same linguistic processes that have been proposed to account for other aspects of inflectional morphology. Indeed, the kinds of correlational features that we have found to statistically predict the form of Australian hypocoristics, such as semantic category and base word length, are exactly the kinds of information sources that connectionist models use to learn regularities and subregularities within linguistic systems. For it is in messy linguistic environments where psycholinguistic models of morphology that exploit phonological and semantic associations between the source (i.e., base) and product (i.e., target morphological form) of hypothesised morphological processes succeed in explaining a greater range of linguistic facts than do item and process models (e.g., Pinker, 1999), which tend to make arbitrary distinctions between so-called rule-based and associative morphological processes.

There is, however, one crucial difference between the morphological processes typically studied by linguists and Australian hypocoristics. Unlike standard morphological processes, Australian hypocoristics do not alter the meaning of the base form. Instead, as an informal register, their use has pragmatic effect. The commonality between Wierzbicka’s (1986) analysis of the –ie and the –o morpheme is that the use of both act as devices that promote social cohesion between speakers. The same could be said for all of the other forms in this class. This is also implicit in other papers that have discussed Australian hypocoristics (e.g., McAndrew, 1992; Sussex, 2004; Taylor, 1992), which point to their frequent use in AusE as deriving from the specific cultural ideals of informality, mateship, and equality. While we agree with these analyses, there is reason to believe that the Australian sociolinguistic landscape has changed sufficiently in the last generation to render this explanation somewhat simplistic. In an as yet unpublished set of experiments (Kidd, Kemp,
Kashima, & Kashima, in prep.), we have investigated Australian people’s attitudes to toward the use of hypocoristics, and have found that the tendency to accept their use interacts with the ethnic background. Specifically, we have found that Australians whose family have been in the country for greater than two generations (i.e., their parents and grandparents were born in the country) are perfectly happy for a speaker to use hypocoristics. In contrast, Australians whose family has a recent history of migration to the country (i.e., parents or grandparents not born in Australia) disapprove of speakers who use them. This suggests that social and ethnic identity has a clear role in the use of hypocoristics, and that speakers of AusE are by no means a cohesive, singular category. Other variables, such as the socio-economic status of the speaker, and accent, are also likely to play a role.

As a final cautionary note, we point out a potential limitation of the methodology we have used in this research. We asked speakers of AusE to provide as many examples of Australian hypocoristics as they could in ten minutes. Therefore, our dataset should only be taken as an indirect measure of hypocoristic use. Regardless of this fact, we argue that it is likely that the examples our participants provided must in some way reflect their usage and the usage of those around them. Analyses of spontaneous language samples of hypocoristic use do not yet exist, but is an obvious direction for future research.
References


### Table 1. Simpson’s (2004) Nine Categories of Hypocoristics.

<table>
<thead>
<tr>
<th>Template</th>
<th>Hypocoristic</th>
<th>Base form</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ie/</td>
<td>Syllable + /i/</td>
<td>coldie</td>
</tr>
<tr>
<td>/o/</td>
<td>Syllable + /o/</td>
<td>journo</td>
</tr>
<tr>
<td>1 syl</td>
<td>One syllable</td>
<td>ump</td>
</tr>
<tr>
<td>/a/</td>
<td>Syllable + /a/</td>
<td>cuppa</td>
</tr>
<tr>
<td>the</td>
<td>‘the’ followed by one or two syllables</td>
<td>The Don</td>
</tr>
<tr>
<td>/as/</td>
<td>Syllable + /as/</td>
<td>chocker(s)</td>
</tr>
<tr>
<td>/s/</td>
<td>Syllables + /s/</td>
<td>mobes</td>
</tr>
<tr>
<td>2 syl</td>
<td>Two syllables</td>
<td>chrysanth</td>
</tr>
<tr>
<td>Acr</td>
<td>Acronym</td>
<td>E</td>
</tr>
</tbody>
</table>
Table 2. Distribution of Hypocoristic Types by Age.

<table>
<thead>
<tr>
<th>Form</th>
<th>Young</th>
<th>Middle-aged</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Rel. %</td>
<td>No.</td>
</tr>
<tr>
<td>/ie/</td>
<td>113</td>
<td>31%</td>
<td>82</td>
</tr>
<tr>
<td>/o/</td>
<td>35</td>
<td>9.6%</td>
<td>27</td>
</tr>
<tr>
<td>1 syl</td>
<td>103</td>
<td>28.2%</td>
<td>44</td>
</tr>
<tr>
<td>/a/</td>
<td>30</td>
<td>8.2%</td>
<td>5</td>
</tr>
<tr>
<td>/as/</td>
<td>2</td>
<td>0.5%</td>
<td>0</td>
</tr>
<tr>
<td>2 syl</td>
<td>43</td>
<td>11.8%</td>
<td>18</td>
</tr>
<tr>
<td>/s/</td>
<td>21</td>
<td>5.8%</td>
<td>7</td>
</tr>
<tr>
<td>Start removed</td>
<td>18</td>
<td>4.9%</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>365</td>
<td>194</td>
<td>167</td>
</tr>
</tbody>
</table>
Table 3. *Type-Token Ratios for each Category by Age.*

<table>
<thead>
<tr>
<th>Form</th>
<th>Young</th>
<th>Middle-aged</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Types/Tokens</td>
<td>Ratio</td>
<td>Types/Tokens</td>
</tr>
<tr>
<td>/ie/</td>
<td>113/292</td>
<td>.39</td>
<td>82/232</td>
</tr>
<tr>
<td>/o/</td>
<td>35/81</td>
<td>.43</td>
<td>27/54</td>
</tr>
<tr>
<td>1 syl</td>
<td>103/187</td>
<td>.55</td>
<td>44/65</td>
</tr>
<tr>
<td>/a/</td>
<td>30/52</td>
<td>.58</td>
<td>5/15</td>
</tr>
<tr>
<td>/as/</td>
<td>2/4</td>
<td>.50</td>
<td>0/0</td>
</tr>
<tr>
<td>2 syl</td>
<td>43/62</td>
<td>.69</td>
<td>18/27</td>
</tr>
<tr>
<td>/s/</td>
<td>21/32</td>
<td>.66</td>
<td>7/11</td>
</tr>
<tr>
<td>Start rem.</td>
<td>18/34</td>
<td>.53</td>
<td>11/16</td>
</tr>
</tbody>
</table>
Figure 1. Example of Bybee’s Lexical Network Model.
Figure 2. Hypothesised lexical network of Australian hypocoristics.