The Words Children Write

2008

Research Summary of the *Oxford Wordlist* Research Study

An investigation of high frequency words in young children's writing and reading development

Professor Joseph Lo Bianco

Janet Scull

Debra Ives

OXFORD
UNIVERSITY PRESS
AUSTRALIA & NEW ZEALAND

100 YEARS OF PUBLISHING
AUSTRALIA 1908-2008

Preface

The *Oxford Wordlist* research study, *An investigation of high frequency words in young children's writing and reading development*, was conducted in Australian schools in 2007. This study is the first of its type in over 30 years.

Wordlists that consider and connect the words used by our broad student community to the classroom increase their relevancy as they are included and referred to in reading, writing and spelling lists. Following repeated requests from educators to be provided with an up-to-date list that reflects Australian students of today, Oxford University Press Australia conducted an extensive and rigorous study to find those words most frequently used by students in their first three years of school in their own writing. The research was designed by Professor Joseph Lo Bianco and Janet Scull from the University of Melbourne, and they have continued their involvement by analyzing the final data.

The Oxford University Press research team was led by Julie Baillie. Julie's work in early years classrooms for over 20 years, and in her role at the school, district and state level working with educators, schools and education systems to drive literacy improvement, was instrumental in the formation of this study.

The *Oxford Wordlist*, the 307 most frequently used words, has been presented as a resource freely available to all Australian educators and this publication, The Words Students Write: Research Summary of the *Oxford Wordlist*, *An investigation of high frequency words in young children's writing and reading development*, is the analysis of this data.

The *Oxford Wordlist* research team

Julie Baillie

Research Manager: Primary Division
Oxford University Press Australia

Julie supported schools, school leaders and educators in all aspects of the process of this study. She collated the data that led to the *Oxford Wordlist*, and the analysis of this data presented in this research summary.

Rosanna Morales

Senior Publishing Editor: Primary Division
Oxford University Press Australia

With a 15 years experience in multimedia design, Rosanna was instrumental in designing the *Oxford Wordlist* interactive data collection tool developed to facilitate the collection of research data and its analysis.

Introduction

While there is extensive research into children's spoken language development, and of their early reading, children's writing appears to be less well documented. Building from this premise Oxford University Press conducted the research reported below, *An investigation of high frequency words in young children's writing and reading development,* and from this produced the *Oxford Wordlist.*

The general aim of the research was to document the words children first write, to examine these choices according to a range of children's demographic characteristics, and to explore what these word choices indicate about children's personal identities and social experiences. This research on the words children write should therefore be understood in the context of the complex and mutually reinforcing growth of children's language: speaking, listening, reading and writing and their educational, social and personal worlds.

What can we learn from the words children use in their earliest samples of writing? What do the words young writers use reveal about children as writers and also about the individual person, his or her subjectivity and interests? What can we infer from the word selections that children make about the influences on them; influences such as gender, social or familial opportunity, ethnic and linguistic background and place of residence? What do children's word choices tell teachers about the teaching of reading and writing? These are some of the many questions raised when conducting and interpreting the *Oxford Wordlist* research.

In the most general, and ideal, terms infants learn language first by learning nouns. This is a process of nominalizing, i.e. naming the world that surrounds them. These nouns that name the world are then qualified, described, given agency, organized and arranged, in other words they are put to work, meeting the child's growing needs for communication. In this way children convert a naming tool into an instrument for social interaction, in turn confirming language as the deepest and most reliable tool a child has available for controlling his or her world. Infants are supported in their early language growth by a veritable cottage industry of targeted communication. Practically every adult engages verbally with infants, tutoring them in their growing language competence, indulging or correcting errors, providing continuous communication. This is a many-to-one profile of communication between intimates and a dependent child. In this scaffolded and attentive environment children develop grammatical accuracy, lexical range and socio-cultural appropriateness.

School involves a radically different communication profile. Here the supportive environment directed at a single child is reversed into a one-to-many communication profile. In classrooms children become multiples, rather than single targets of directed communication, they become 'students' and enter formal relationships with adults who are no longer intimates but professionals *in loco parentis*. The single source of communicative input, the teacher, must divide his or her time among many learners.

In schooling the astonishingly complex process of language acquisition is further extended into literacy. Speaking and listening are connected through meaningful interaction and reading and writing elaborate these primary language processes into formal routines of knowledge acquisition. The speaking/listening pair is complemented by the reading/writing pair with one being mostly productive and the other mostly receptive. Schools therefore extend the primary

language socialisation of the home and induct children into the two literate equivalents of home communication: reading and writing. Reading, like listening, involves absorbing the messages of others; writing, like speaking, involves using the most elaborate manifestation of language to communicate messages to others.

As young children learn to read they draw heavily on their spoken language repertoire, as they learn to write they draw heavily on their reading skills. There is a close relation between the particular language skill being acquired and the one already mastered, i.e. the range and nature of each skill that children have under control informs and stimulates subsequent skill development, in an organic, cumulative and mutually reinforcing way. Both of the *receptive* skills (what they understand of the language of others, especially teachers, parents and peers) are analogues of their *productive* skills (what they produce, in speech, proto-writing and early writing). Reading is a source of language input and writing is a form of language output; each skill grows in mutual relation to the other.

In the beginning stages children read words that most easily fall within their spoken language experiences and this holds true of what they write. Over time the strict interdependency between speech and writing, along with the relation between concrete and abstract ideas, diverges, and so too does the curriculum of schools. Each involves tools that serve to mediate the experience: pens, paper, keyboards and ever more integrated digital tools. All encapsulate the expanding *codes* that store symbolic information and are delivered through diverse *modes* of communication and come to represent ever more sophisticated *meanings* (Lo Bianco and Freebody, 2001).

Research methodology and design

The principal aim of the *Oxford Wordlist* research was to gather a sufficiently large and representative sample of children's 'free' writing to facilitate differentiated analysis of some key characteristics. As a result, the methodological challenge was to generate a sample sufficiently large to permit reliable analysis of both overall and sub-group characteristics, and to elicit the sample texts in asnaturalistic a way as possible. Moreover, it was essential to tap into children's authentic writing in such a way as to uncover both what is common to all young writers and what differentiations exist.

The research was intended to be cross-sectional, rather than longitudinal, and so few comments are made about dynamic aspects of children's word usage, other than references to what can be gleaned from comparison with previous research on children's early written words. The cross-sectional, point-in-time sampling permits the researchers to observe contemporary Australian English in use in the most de-contextualized and elaborate of the four main language skills.

It was decided to restrict the sample to the writing practices of children in the first three years of school. The main demographic differentiations investigated were gender, socio-economic status, language background and location. Feasibility constraints determined population size and spread, so that it was decided to gather writing samples from 1000 children in two states, Victoria and South Australia. Relevant processes were followed to gain permission from the governing educational authorities allowing researchers access to schools. Proportions for each of the selected demographic factors were calculated using the Australian Census figures (ABS, 2007), enabling participant numbers in each of these cohorts to be broadly reflective of its representation in the whole Australian population.

Five writing samples were collected from participating children in the first year of school and three samples from children in the second and third years of school. The reason for collecting five writing samples from first year students was to provide a similar overall total word count to that which would be yielded from the samples collected from samples in years 2 and 3. To preserve as naturalistic a process as possible, sample collection was allocated to classroom teachers who were requested to conserve texts produced by their students during regular writing sessions. In this way, the children were also free to use the writing tools and aids regularly available in their classrooms and this facilitated their use of the words they *wanted* to write rather than being limited to those words they *could* or *were asked to* write. While acknowledging that the classroom focus on specific topics might influence the content or choice of text type and hence the grammatical structure of the collected texts, the condition of instruction to teachers was that they should collect samples of 'undirected' or 'free' writing. The aim was to ensure that the sample would tap into children's relatively unconstrained expression and provide evidence of personally preferred word choices, hypothesizing that this would throw light on the personal lives, skills and interests of the children involved.

The following demographic data is from the 3776 samples included in the analysis. The figures in Table 1 refer to the number of useable texts collected, followed by the total word count represented by these texts, so that 1879 texts were collected from boys comprising 73,430 words, and 1897 texts from girls comprising 86,808 words. Variations from the totals are indicated. Socio-economic status (SES) was not determined directly but on the basis of the location of the school and its SES classification within the relevant education department procedures. All cohort factors, such as language and cultural background, were determined by teachers according to school records reported to departments of education.

Gender

Male	Female
1879 (73,430)	1897 (86,808)

School year

First year	Second year	Third year
1891 (38,733)	951 (48,798)	934 (72,702)

School setting (SES) NB n=3771**

Low SES	Mid SES	High SES
1450 (62,151)	698 (25,836)	1623 (72,136)

Indigenous NB n=3765**

Indigenous	Non-indigenous
63 (902)	3702 (157,483)

Indigenous = 1.67% of all

Language

English speaking background (ESB)	Non-English speaking background (Non-ESB)
2466 (97,616)	1310 (62,621)

Non-ESB = 34.69% of all

Location

Urban	Rural
3321 (144,274)	455 (15,964)

Urban boys	Urban girls	Rural boys	Rural girls
1624 (66,019)	1697 (78,255)	255 (7,411)	200 (8,553)

Table 1: Demographic data

The collected samples were copied and student codes were attached to each sample along with demographic data completed by the class teacher. Only complete data sets for each child were retained for analysis. This process produced a total of 3776 texts distributed as presented in Table 1. The total sample was then subjected to word count and differentiated analysis according to the cohorts in the sample. Proper nouns and brand names were deleted from the word count.

Research results

What's common among early writers?

The most striking and general observation from the data is the high proportion of commonality among early writers. Regardless of their location, gender, ethnicity, or socio-economic differences Australian children write with a large common vocabulary, they write words their peers will understand and therefore they represent a 'discourse' community, or a community of communication.

The imaginative, descriptive and declarative worlds these children deploy in what they write indicates a shared enculturation, a common world of Australian childhood, which suggests that despite other differences Australia's young writers inhabit an overlapping world of experience that they display in their writing. This shared vocabulary is evident at two cut-off levels: the top 20 and the top 50 words. It is only beyond these thresholds that indicators of difference begin to emerge. Differences of gender and probable differences in socio-cultural experience, as well as the personal imaginative and creative capabilities of individual children, become more apparent when the top 100, 200 and 300 words are compared. This suggests that the personal subjectivity of individual children is displayed and diverges from the shared community of expression once the essential core of vocabulary permits meaningful differences to emerge.

A significant point of note is that on 16 occasions when a verb appears in the 307 words in both its past and present tense forms, the past tense verb appears higher in the frequency order than its paired present tense verb. 18 additional past tense verbs were also featured in the 307 words whereas only 6 additional simple present tense verbs make the top 307 words. As 52% of the text types written by all students in the study were recounts this prevalence of past tense forms reflects the dominant grammatical structure of recount texts.

Other features at this broad level are children's frequent use of the apostrophe to signal word contractions. In the top 307 words used by all children the contractions <*didn't*>, <*it's*>, <*don't*>, <*I'm*>, <*couldn't*> and <*that's*> are used, perhaps showing how children's oral language articulates into their writing. An apostrophe indicating possession becomes apparent in children's use of <*cousin's*>, a feature that is particularly prevalent in the writing samples collected from children of non-English speaking backgrounds.

Generation change

Despite differences in design and size the *Oxford Wordlist*, at this aggregated level, can be usefully compared to the last known Australian study of words in children's writing, the Salisbury Word List. This comparison reveals interesting changes over time.

The Salisbury Word List was compiled in Adelaide, South Australia on research conducted during 1978–79 and records the 2000 words most frequently encountered in children's writing, based on 150 text samples collected for each year level between years 3 and 7 (Education Department of South Australia, 1984). As such the sample is considerably smaller and more restricted than the data in the *Oxford Wordlist*, and also differs in its spread over a different span of school years. The 30 years of difference are significant generationally. During the period from 1978 to now writing itself has also undergone profound change as a result of the introduction and diffusion of digital communications and other devices of information technologies that have 'technologised' (Cope and Kalantzis, 2000; Snyder, 2002; Lankshear and Knobel, 2003) what counts as literacy and shaped some of the processes whereby writing is produced.

Table 2 shows those words included in the top 200 words from the *Oxford Wordlist* that do not appear in the top 200 words from the Salisbury Word List. It is interesting to observe that although there is 72.5% agreement across the two lists 55 words differ. For ease of reference the words are listed alphabetically.

also	eat	games	party	swimming
ate	ever	it's	played	TV
baby	fairy	liked	playing	upon
ball	family	lost	really	want
best	favourite	lots	Saturday	wanted
bike	fish	love	scared	was
birthday	food	movie	shop	watched
bought	football	named	shops	weekend
brother	Friday	new	sister	woke
cake	fun	outside	stayed	won
cousin	game	park	Sunday	would
dinner				

Table 2: Oxford *Wordlist* words from the top 200 that do not appear in the Salisbury Word List top 200 words

The most striking feature of the differences between the two lists relates to the greater detail that Australian children today devote to events, activities and relationships than their counterparts of 1978–79. There is also the intriguing presence of a vocabulary of consumerism and a greater interest in food. The words <*bought*>, <*new*>, <*shop*>, <*shops*>, <*want*>, <*wanted*>, <*eat*>, <*ate*>, <*cake*>, <*dinner*>, <*fish*>, <*food*> do not appear among the top 200 Salisbury Word List words. Interest in food and consumer activities continues to be a feature in the *Oxford Wordlist* with the appearance of <*lunch*>, <*breakfast*>, <*chips*>, <*icecream*>, <*shopping*> and <*buy*> in the subsequent 100 words.

There is also a notable extension of the range of words used to describe viewing, so that in 2007 children write <*see*>, <*saw*>, <*watched*>, <*movie*>, <*TV*> compared to the more limited range of <*see*> and <*saw*> in 1979. The two lists also provide insight into the expanded leisure activities of children. In today's writing Australian children account for activities involving the following words: <*ball*>, <*bike*>, <*football*>, <*park*>, <*played*>, <*playing*>, <*soccer*>, <*swimming*>, <*birthday*>, <*party*>, <*fun*>, <*game*>, <*games*>, <*won*>, <*Saturday*>,

<*Sunday*> and <*weekend*>. The use of the words <*game*> and <*games*> could also be used to describe viewing when attributed to activities related to interactive console game use.

There is the probability of a deep and subtle change in sense of what constitutes 'family' across these young writers separated by 30 years. The *Oxford Wordlist* excludes references to <*mother*> and <*father*> from the top 200 words but it does share with the Salisbury Word List a consistent use of <*mum*> and <*dad*>. Discarding <*mother*> and <*father*> from the top 200 goes along with the diminished use of the titles <*Mr*> and <*Mrs*> reinforcing the possibility of a more general decline in expected formality, at least at the referential level.

Comparison with the *Oxford Wordlist*, across the differentiated cohorts, reveals further preferences and interests across the sample sets, as discussed in the following sections.

Gender

Similarities, and persistent and significant differences, characterize the word frequencies of boys and girls.

In the top 100 words, boys and girls commonly write of things they have <*watched*>, they report temporally about the <*weekend*>, and separately specify <*Saturday*> and <*Sunday*>, and they nominate places, particularly <*park*>, and events, especially <*birthday*>. This homogenous written-about social world of early writers does begin to show internal differentiation when we look to how these activities, places, times and events are described, accounted for and evaluated. Boys more frequently write about a <*game*> they have <*won*> and list items they have <*found*> whereas girls write about things they have <*bought*> and about circles of intimacy, expressed mostly as <*friend*>, <*girl*> and <*sister*>.

What is especially revealing in the sample, however, is what we can infer by what is absent as much as by what is present. Sports are not listed among the overall boys' top 100 words, and this suggests that the use of <*won*> and <*game*> are associated instead with computer and interactive console games.

In many of these games players search for tools, weapons and treasure in order to prevail in the competition set up by the game or to gain assistance in their quest. The absence of <*won*> and <*game*> among the girls' top 100 words seems to suggest less prominence of computer and interactive console game playing, and perhaps underscores a gendered pattern of participation in computer-mediated play.

Girls make use of a wider range of possessive adjectives, i.e. <*her*>, <*our*>, <*their*> in addition to the shared use of <*his*> and <*my*>. This appears to suggest a greater differentiation among girls of the language of interpersonal relationships, so that they allocate more detailed consideration to accounting for and describing how people relate to each other.

Interesting referential data is revealed in the top 300 words used by girls and boys. At this level of analysis it becomes clear that while boys and girls both use <*boy*> it falls only to girls to refer to <*girls/girl*>. In the top 300 category <*he*> is used 1.8 times more frequently than <*she*>; while <*him*> and <*his*> are 1.45 times more commonly encountered than <*her*>. This too reinforces a sense that young female writers name a world of social relationships with more detail and sensitivity than their male counterparts.

Girls also used a wider range of gendered nouns such as <*aunty*>, <*fairy*>, <*girl*>, <*girls*>, <*prince*> and <*princess*> and included in their writing, in what is apparently a generic category of fantasy, the word <*witch*>, never mentioned by boys. The only similar noun used exclusively by boys is <*king*>. Gendered experience is also marked by pronominal reference, so that boys referred to male interactants significantly more often than to female ones: <*he*> registers 1349 times among the boys while <*she*> registers only 155 times.

By contrast, while girls refer to female interactants more often than to male ones, the disparity in their sample is far less pronounced than in the boys' texts. Girls showed more equitable usage of these pronouns with <*he*> occurring 621 times and <*she*> 910 times. Table 3 shows this use of gendered words used by girls and boys across their top 300 words.

Word	Boys: Frequency	Girls: Frequency	Word	Boys: Frequency	Girls: Frequency
aunty	0	36	his	370	187
boy	107	67	king	31	0
brother	135	159	man	81	48
dad	322	453	mum	288	504
fairy	0	108	prince	0	40
girl	0	154	princess	0	62
girls	0	43	she	155	910
he	1349	621	sister	69	166
her	72	529	witch	0	45
him	188	126			

Table 3: Children's use of gendered words

Language background

Language background is a missing category in previous research. The *Oxford Wordlist* research collected a substantial sample from language backgrounds other than English.

Given the small individual group numbers from a range of language sets, the analysis here proceeds according to a broader distinction between children of English speaking home backgrounds (ESB) and those from homes where English is not the main language. Even this distinction, however, requires caution, given that few non-ESB home environments are monolingual, and so the categories are idealized rather than watertight. Unfortunately the non-ESB group does not differentiate children who are learning English as a second language, who are instead subsumed within the broader cohort of children who come from homes in which one or both parents speak a language other than English or who were born in a non-English speaking country.

Despite these limitations the data reveal interesting similarities and differences. Among the top 100 words used by these children there is 90% agreement across both groups, meaning that 90% of the words are identical, decreasing to 85.5% agreement in the top 200 words. This is sufficient to justify the distinction and is revealing about the probable impact of language background on word preferences.

Both cohorts write of playing games and watching TV, but a number of other activities appear to distinguish between them. Children from home backgrounds where only English is used

include the words <*beach*>, <*bike*>, <*boat*>, <*football*>, <*soccer*>, <*swimming*> and <*zoo*> in their free writing, whereas children from non-English home backgrounds write less often of active pursuits and feature the following words more prominently in their writing: <*book*>, <*movie*>, <*garden*>, <*computer*> and playing with <*cousins*>.

Children from homes where English is not the main, dominant, or sole language, more frequently write about family events: choosing words like <*ate*> and <*family*>. These give the flavour of the social occasions involved, and link to other references associated with extended family networks and sociality, implied by the prevalence of <*cousin*>, <*cousins*> and <*cousin's*> in writing samples collected mainly in the third year of school. <*Cousin*> and <*family*> are not listed among the top 100 words for children from homes where English is the dominant language, and <*ate*> lessens in frequency after the second year of school. The persistence of references to <*ate*> by non-ESB children is notable.

Location

The writing samples of rural students represent 13.7% of the urban entries and provide a total of 15,964 words compared to the 144,274 words written by their city counterparts. Nevertheless, a comparison of the urban and rural sample is instructive in some predictable and also possibly some surprising ways. The data show some interesting word choices that might be particular to location and to population density or frequency of encountering strangers.

The *Oxford Wordlist* research analysis indicated that while rural children often write about the <*weekend*> they do not specifically refer to <*Saturday*> and <*Sunday*>. This also holds true when the data are disaggregated to investigate the words used by rural boys and girls, possibly indicating that country children experience fewer structured or routine weekend activities than their urban peers. In respect to consumption there is a markedly less frequent use of <*bought*> (only six occurrences) by rural children and the absence of <*ate*>. The absence of these from the full list of 300 words probably reflects greater distances from stores and entertainment venues and therefore shopping, consumerism and organized entertainment. Perhaps similarly, while urban girls write about a <*party*>, this is not an event commonly referred to by rural girls.

Gender differences also mark the sample of rural children in particular ways. <*Motorbike*> appears as one of the top 100 words used by rural children. The word makes 34 appearances in the rural boys samples, but is totally absent from the words rural girls write. Girls instead refer to <*horse*> (13 occurrences). Interestingly, <*motorbike*> is used far more frequently than <*bike*> (15 occurrences) by rural boys and, possibly surprisingly, <*bike*> does not feature at all among rural girls' top 100 words. Moreover, <*watched*> is ranked at number 108 (with 9 occurrences) on rural boys' count of words, but it is ranked 89 for rural girls who used it 14 times across their writing samples. The difference is not stark but could possibly imply more sedentary activities, depending on the context in which the word <*watched*> is used in their writing.

Interestingly, the binary paired relationships <*brother/sister*> and <*friend/friends*> do not appear among the top 100 words used by rural children. It is possible that rural children might be using proper nouns to identify individuals in these sibling or friendship roles.

These data are striking in some respects, and while this and the preceding rural/urban observations may be significant, the disparity in small sample size between the urban/rural categories reduces their generalizability and they must remain observations.

School year level

School year level has a moderate influence on the words commonly used by children in the first year with 13% of words differing from those frequently used by all children. The wordlist data collected in this research indicate that the younger children use a large number of nouns in their writing while the older children appear increasingly able to manipulate a wider range of grammatical forms in order to vary sentence structure, expressive content and general nuance of meaning.

The subjective nature of nouns leads to the increased variation that is consistently evident in the wordlists attributed to children in the first year of school. Six of the words used by these young children are nouns that reflect their interests and activities, i.e. <*football*>, <*party*>, <*shop*>, <*shops*>, <*toy*> and <*zoo*>; and five are verbs. Of the latter, two convey a clear action (i.e. <*gave*>, <*see*>) while two are sensing verbs (i.e. <*liked*>, <*love*>). This contrasts with the high level of commonality (94%) evident in lists compiled from students' writing in the second and third years of school. However, gender appears to have a greater influence than school year level.

Words used by boys and girls in their first year of school

There are many similarities in the words used by children in the first year of school. Family, friends and computers figure prominently in the words both boys and girls write. However, there are also some stereotypical differences. There are some striking differences in boys' and girls' lists about imputed gender related to the animals, sports and the three Ms of fantasy writing – the mythical, the magical and the monarchical – all of which figure prominently in their writing.

While boys and girls frequently write about dogs, boys often bring into their texts 'dangerous' animals such as the <*snake*>, <*shark*>, <*lion*>, <*tiger*> and <*crocodile*>, although the <*octopus*> may gain its place because of strangeness rather than implied danger. In this context it is possibly significant that the word <*teeth*> appears (at number 298) on the boys' wordlist but is absent from the words girls use. In girls' writing a different menagerie comes into play. In girls' texts the animal kingdom is represented by the <*cat*>, <*horse*>, <*zebra*>, <*monkey*>, <*giraffe*> and <*puppy*>. Dangerous animals are not absent from the list of young female writers, who include <*snakes*> and <*sharks*> in their stories, but they are referred to less frequently. In general girls introduce fewer dangerous animals, and more domesticated, or placid ones, into their stories and, as <*puppy*> suggests, prefer to name them in ways that elevate intimacy.

Both boys and girls often write about <*football*>, but with some revealing differences in reference forms and morphology. Boys alternate their references to football with <*footy*> (also used to label the actual ball) and discuss <*soccer*>. While boys also write about <*cricket*>, the only other team sport mentioned by girls is <*basketball*>.

The mythical, magical and monarchical parade through the children's texts in the first year of school, presumably re-circulated in video and computer iconography as well

as in traditional reading. Both boys and girls write of <*magic*> and <*castles*>, but are differentiated by exclusive use of the following nouns. Boys make reference to <*dinosaur*>, <*dragon*>, <*monster*>, <*king*>, <*ghosts*>, <*pirate*> and <*robot*>, whereas in the writing of girls the words <*fairy*>, <*fairies*>, <*mermaid*>, <*princess*> and <*witch*> are found.

These differences reside in the imaginative realm; even if writing genres, topics or tasks are scripted by teachers, consistent and predictable domains of interest appear to stem from children's early socialization.

Words used boys and girls in the second year of school

A number of the gender differences noted in the first year of school had lessened by the second year. In particular, the list of animals is not as clearly differentiated between girls and boys and in the second year girls start to add <*soccer*>, <*prince*> and <*monster*> to their stories. However, if this is a convergence of interests it would seem that boys do not reciprocate, at least not in some key areas, preferring not to include <*fairies*> and <*princesses*> in their texts. Both boys and girls continue to refer to <*computer*> more frequently than to <*TV*>. There are also some persisting differences in the words used to discuss natural phenomena and the environment. In boys' writing <*fire*>, <*world*>, <*earth*> and <*cave*> feature prominently, conjuring up a world of primordial categories, whereas for girls <*water*>, <*moon*> and <*forest*> distinguish their writing.

The strongest differentiator is the boys' use of the words <*fight*> and <gun>. Analysis of the most frequently used verbs places these in close proximity to <*killed*>, <*shot*> and <*kill*>. On its own the *Oxford Wordlist* research cannot supply the essential contextual information required to interpret the preponderance of such words. An examination of the texts themselves identified the true nature of this kind of writing, and the persistence of words that denote action, struggle, mortal danger, and indeed mortality itself in boys' free writing poses interesting questions for educators.

Word used by boys and girls in the third year of school

There are high levels of agreement between words written by students in the second and third years of school. Close analysis of these two lists provides insight into children's developing interests as changes in the words used show that <*computer*> is no longer a high frequency word for girls and <*TV*> climbs above <*computer*> in frequency of references by boys.

There are also differences with regard to words used to describe sport with boys often making recourse to <*goalls*>, <*match*> and <*team*>. Additionally, while boys have intensified their focus on <*soccer*> and <*football*> in that order, for girls it is <*game*> that may indicate sporting vocabulary, though it is possible that <*game*> could have a wider frame of reference, such as playing a <*game*> on the computer or interactive games console.

Words used to refer to the three Ms persist, but remain differentiated, with boys continuing to use <*dragon*>, <*ghost*>, <*king*>, <*monster*> and <*dinosaur*> while girls remain loyal to <*fairy*>, <*princess*>, <*witch*>, <*prince*> and <*beast*>.

Also retaining its hold in boys' writing is <*fight*> but by the third year of school it is joined by <*battle*>. This kind of action prose remains linked to <*killed*>, <*shot*> and <*dead*> on the full wordlist for boys. These words are for the most part missing from the girls' wordlist.

Socio-economic status (SES)

Comparisons of the top 100 words used by all students with those of low, mid or high SES ranges show very high levels of correspondence (between 91% and 93% agreement). This suggests that SES, in the broad way in which it was included in the research, is not a strong predictor of the 100 highest frequency writing words. However, while this remains true at the broadest level of comparison, possibly significant differences can be discerned when factors such as gender or year level are linked to the SES category.

As the mid SES sample contains a lower number of writing samples than either of the other subgroups – 204 entries compared to 363 (low SES) and 384 (high SES) – the focus for differentiation of gender and SES is the low and high range only as they have similar entry numbers and therefore are not likely to show the degree of variation commonly noted with smaller sample sizes.

When examining the top 300 words differentiated into gender and linked to low or high SES ranges some significant differences are noted. The level of correspondence of words for girls of the low SES group and girls of the high SES girl is 83%. The words used by girls in the low SES sample but not by girls in the high SES sample are heavily linked to functional activities such as attending school. These words include <*teacher*>, <*Miss*>, <*read*>, <*class*>, <*work*>, <*book*>, <*bag*>, <*learnt*>, <*Monday*>, <*Friday*>, <*kids*> and those words indicating familial relationships such as <*aunty*>, <*cousins*>, <*cousin's*>, <*mum's*>, <*dad's*>. Only the girls from the high SES group refer to the following animals: <*butterfly*>, <*duck*>, <*frog*>, <*rabbit*>, <*shark*>, <*snake*>, <*animals*> and <*horse*>. These animals are often featured as the main subject of a narrative text. Words that denote more leisure time activities such as <*beach*>, <*boat*>, <*holiday*> and <*movies*> also feature only on the high SES girls list. Two sporting vocabulary words <*soccer*> and <*team*> appear on the high SES girls list and are not used by the girls of the low SES group in the sample.

Similar results are evident when linking boys with SES. There is a much lower correlation between the words on the full list of 300 words from the sample of high SES boys and the 300 words from the sample of low SES boys (78%). Words used by boys of low SES but not by boys of high SES include words used to describe shared activities and events with family members: <*barbecue*>, <*breakfast*>, <*cousin*>, <*cousin's*>, <*cousins*>, <*dad's*>, <*eating*>, <*family*>, <*everyone*>, <*icecream*>, <*kids*>, <*pizza*>, <*sleep*> and <*children*>. An examination of the writing samples of this group reveals that a significantly higher level of recount texts are written (62% compared to 46%). This could also possibly explain the use of time connectives by low SES boys not used by the high SES boys group. These include <*yesterday*>, <*Wednesday*>, <*Monday*> and <*today*>. Just as significantly a much higher percentage of narrative texts make up the samples of the high SES boys group when compared to the low SES boys group (37% compared to 17%). This may explain the inclusion of the words <*castle*>, <*battle*>, <*king*>, <*land*>, <*ghost*>, <*happily*>, <*monster*>, <*suddenly*> and <*died*>. These words are associated with the imaginative realms of the mythical, monarchical and magical worlds.

Indigenous children

The overall sample contains a small data set provided by Indigenous children, the low numbers restricting the research; however the following observations can be made.

The Indigenous children who participated in this study did not frequently write about <*mum*>, <*home*> or <*school*>. They did however refer to <*dad*> and <*house*>. Additionally, while these children write about the <*weekend*> (as do rural children) they do not often explicitly write about <*Saturday*> (word 213 with a frequency of 1) and <*Sunday*> was absent from the list. This is especially interesting because only three of the 63 Indigenous entries were written by children living in rural settings.

A far higher number than for other groups of words commonly used by Indigenous children are activity based, i.e. <*play*>, <*played*>, <*playing*>, <*band*>, <*basketball*>, <*beach*>, <*bike*> and <*football*>. Although, in part this might be an artefact of the small sample size and the related predominance of nouns that will appear with relatively lower frequency for larger cohorts, it is still useful to note that while non-Indigenous students refer to <*play*> and <*played*>, the names of the associated sports and games do not appear among their top 70 words.

Analysis of the text types written by Indigenous children (i.e. recount, description and narrative respectively) reveals further patterns in word usage. A total of 32 of the 63 entries were recount texts (word count of 462) and the top 20 words used in this text type indicate that these children are often writing:

- <On the weekend I/we went … >
- <On the weekend, I/we played … >
- <On the weekend, I/we watched …>
- <It was fun.>
- <I/we had fun.>.

Implications for educators

The lack of similar studies and the long gap between the previous collection of this kind of data suggest possibilities and challenges for teaching reading, understanding and teaching writing and promoting discussion about the inter-related processes of language acquisition.

Detailed analysis of these data suggests three main observations on the current research.

1. The first and dominant impression is the high degree of consistency within the first 100 words for all groups. Beyond expected similarities in the use of 'function' or 'tool' words (i.e. adverbs, prepositions, conjunctions, adjectives and verbs) this pattern of overall commonality in lexical expression underscores a shared cultural experience for young Australian children that is independent of their gender, their socio-economic status, cultural and language backgrounds and geographic location. When Australian children write stories and convey images and messages about their lives, experiences and values they are speaking more than a common language; in fact they are expressing a shared communal and national culture.

2. Alongside this coherence of experience and culture there is also interesting, and significant, divergence, as the various cohorts particularize the shared verbal culture into a range of differentiations reflecting the sub-categories of gender, location, social class, cultural background and home language background. While not all of these differentiations are equally significant, gender being the clearest and most persistent marker of imaginative difference as expressed in word choice in writing, they serve to give concrete evidence of how difference articulates or emerges from similarity.
3. The third broad observation relates to the generational shift that can be identified between the present data set and its predecessors. The Salisbury Word List top 200 words and the *Oxford Wordlist* top 200 words share 72.5% agreement in word choice noted among the young writers. A 27.5% divergence must be considered very substantial given that a gap of only 30 years separates the two data sets.

The *Oxford Wordlist* represents contemporary Australian English and the non-random, systematic patterning of the word differences between the two studies exposes important changes in the socialization of young people and the impact of technology on writing and social activity. Moreover the *Oxford Wordlist* draws on a differentiated sample of learners, which is sufficiently representative of the range of young learners in Australian schools.

Comparison with other wordlists

How do the Australian research studies, and in particular the *Oxford Wordlist*, compare with other English language research of this kind? In order to respond to this question, it is important to emphasize that the *Oxford Wordlist* differs fundamentally from many other lists in its collection methods. It is based on children's usage, words they know orally and visually, rather than being derived from a study of words in children's reading texts. In this way, the *Oxford Wordlist* differs from a recent United Kingdom database of words gathered from children's reading texts, as reported by Stuart, Dixon, Masterson and Gray (2003).

In common with *Oxford Wordlist* research, Huxford, McGonagle and Warren (1997) examined changes in high frequency wordlists over time by comparing the words used in 1254 free writing samples collected from 4, 5 and 6 year olds with three lists in common usage in the United Kingdom, namely those of Edwards and Gibbon (1964), Fry (1980) and Reid (1989). These writers found total agreement on more than 50% of the words from the comparison lists. It is also interesting to note that there is a 75% agreement between the Huxford et al's top 100 words used by English children, representing research conducted in the late 1990s in the United Kingdom, and the present *Oxford Wordlist* of words written by Australian children in 2007.

The most cited of all wordlists is however the Dolch wordlist of basic sight vocabulary (1936). This was derived from a compilation of children's oral vocabulary and words commonly found in young children's reading materials. Although these data were collected in the 1920s and early 1930s and a comparison of the 220 words of this list with the top 220 words on the *Oxford Wordlist* reveals an agreement level of only 54%, this not only reflects changes in language usage over time, it is also related to Dolch's decision to exclude all nouns. 60 of the top 220 words on the *Oxford Wordlist* are clearly identifiable as nouns.

The *Oxford Wordlist* has been prepared from a database far larger and more representative than any Australian and most international predecessors, using text samples gathered naturalistically from free writing practice in regular classroom operations. It seeks to respond to all the above needs: to inform teachers of the research base behind the development of texts, to assist in informed text selection, to record and analyze changes and patterns of children's word use in writing for the whole population and a group of sub-population categories, and to allow teachers to expand children's writing language repertoire by comparing and contrasting an individual learner's performance with the words used by their peers.

The research, as we have noted, also offers insight into patterns of culture across Australia, so that changes in children's lives are reflected in what children choose to say to the audiences to which they direct their written words.

Words children choose to write

Luke (1994) holds that teachers need to understand "how texts and literate competence fit into larger patterns of community and social structure" (p. 4). He states, drawing on Foucault, that schooling "involves inculcation in a series of validated discourses, texts and genres" (p. 12) of language. We have seen from the data supplied above how boys and girls, ESB and non-ESB, rural and urban children engage in processes of inclusion and exclusion from intimate, personal and familial networks from their earliest samples of writing. The validation may simply be the immediacy of experience that young Australian children gain from contexts in which they are immersed, intimate peers, <*cousins*>, imagined heroic roles as well as changing family patterns in which Mr, Mrs, Father and Mother have greatly declined as our culture encourages a more egalitarian ethos in relationships. Today's young children interact horizontally in virtual worlds which introduce into their lives, and which appear in what they choose to write, words (and presumably over time also discourses, texts and genres) that will impact on the language used in classrooms.

The *Oxford Wordlist* and the research on which it is based aims to enhance the quality, interest and awareness of text production *for* children based on text production *by* children. It is not only reading, but also writing capability that is important for children in achieving social and intellectual outcomes from formal schooling that can enable them to lead meaningful and productive lives. Luke has argued concerning 'evidence-based approaches' to literacy education policy that improved literacy standards mean that schools ought to be assisted to identify and connect with the "kinds of capital and knowledge, cultural and linguistic resources" that are valued in the community (2005, p. 670).

These are the aspirations which have prompted the current research to examine children's word usage in undirected writing, intent on reflecting the current Australian language and to capture something of the breadth of experience various categories of children bring to their engagement with the highest and most abstracted manifestation of language: writing, and therefore <*words*>.

References

Australian Bureau of Statistics (ABS). (2007, 24 January). *Year Book Australia, 2007*. Retrieved 7 May, 2007, http://www.abs.gov.au/ausstats/abs@.nsf/mf/1301.0.

Baker, C., and Davies, B. (1993). Literacy and gender in early childhood. In A. Luke and P. Gilbert (Eds.), *Literacy in contexts: Australian perspectives and issues* (pp. 55-67). St Leonards, NSW: Allen and Unwin.

Baker, C., and Freebody, P. (1989). Children's first school books: introductions to the culture of literacy. Oxford: Blackwell.

Brabham, E., and Villaume, S. (2002). Leveled text: the good news and the bad news. *The Reading Teacher*, 55(5), 438-441.

Clay, M. M. (1991). Becoming literate: the construction of inner control. Portsmouth, NH: Heinemann.

Clay, M. M. (2001). Change over time in children's literacy development, Auckland:Heinemann.

Cope, B., and Kalantzis, M. (eds), (2000). *Multiliteracies*. London: Routledge.

Dolch, E. W. (1936). A basic sight vocabulary. *The Elementary School Journal*, 36(6), 456–460.

Education Department of South Australia. (1984). Salisbury word list. In Education Department of South Australia (Ed.), *Spelling R–7 Language Arts* (pp. 73–80). Adelaide: Education Department of South Australia.

Edwards, R. P. A., & Gibbon, V. (1964). *Words your children use*. London: Burke Publishing Company.

Fountas, I. C., and Pinnell, G. S. (1999). *Matching books to readers: using leveled texts in guided reading, K-3*. Portsmouth, NH: Heinemann.

Fry, E. (1980). The new instant word list. *The Reading Teacher*, 34(3), pp. 284–289.

Goodman, K. S. (1971). Reading: a psycholinguistic guessing game. In H. Singer and R. B. Ruddell (Eds.), *Theoretical models and processes of reading*, Newark, DE: International Reading Association.

Hiebert, E. H., and Fisher, C. W. (2007). Critical word factor in texts for beginning readers. *The Journal of Educational Research*, 101(1), pp. 3–11.

Hill, S. (2001). Questioning text levels. *The Australian Journal of Language and Literacy,* 24(1), pp. 8–20.

Huxford, L., McGonagle, R., and Warren, S. (1997). Which words? Words which 4 to 6 year old children use in their writing. *Reading*, 31(3), pp. 16–21.

Lankshear, C and Knobel, M (2003), *New literacies*, Open University Press: Buckingham.

Lo Bianco, J. and Freebody, P. (2001). *Australian Literacies: Informing National Policy on Literacy Education*. Melbourne: Language Australian Publications.

Luke, A. (1994). *The social construction of literacy in the primary school*. Macmillan Education Australia: South Melbourne, Victoria.

Luke, A. (1999, October 20). *Education 2010 and new times: why equity and social justice still matter*. Retrieved February 22, 2008, from http://education.qld.gov.au/corporate/newbasics/docs/onlineal.doc.

Luke, A. (2005). Evidence-based state literacy policy: a critical perspective. In N. Bacia, A. Cumming, A. Datnow, K. Leithwood and D. Livingstone (Eds.), *International handbook of educational policy* (Vol. 2, pp. 661–675). Dordrecht, The Netherlands: Springer.

Olson, D. R. (1994). *The world on paper*. Cambridge University Press: Cambridge.

Reid, D. (1989). *Word for word*. Wisbech, Learning Development Aids: UK.

Rog, L. J., and Burton, W. (2001). Matching texts and readers: leveling early reading materials for assessment and instruction. *The Reading Teacher*, 55(4), pp. 348–356.

Snow, C. E., Burns, M. S. and Griffin, P. (Eds.) (1998). P*reventing reading difficulties in young children*, National Academy Press: Washington, DC..

Snyder, I. (Ed.) (2002). *Silicon literacies*, Routledge, London.

Stuart, M., Dixon, M., Masterson, J., and Gray, B. (2003). Children's early reading vocabulary: description and word frequency lists. *British Journal of Educational Psychology*, 73(4), pp. 585–598.

Professor Joseph Lo Bianco (AM, FAHA)

Chair of Language and Literacy Education
Associate Dean (International)
The University of Melbourne

Professor Lo Bianco is an internationally renowned researcher and author, particularly in the area of English as a second language and language planning. He is best known as the author of the 1987 National Policy on Languages, adopted as a bipartisan national plan for English, Indigenous languages, Asian and European languages, and Interpreting and Translating services and now used in many international locations as a model of rational language planning. He has an extensive publications record in literacy education, multiliteracies, Asian studies, bilingualism and bilingual education, language policy and culture in second language teaching. Professor Lo Bianco has more than 100 scholarly refereed articles and has written or edited 25 books and reports.

Janet Scull

Lecturer
Language and Literacy Education
The University of Melbourne

Janet Scull is an experienced classroom teacher and as a key author of the Victorian Early Years Literary Program has contributed to design of systemic approaches to literacy teaching and learning. Janet is also a Reading Recovery Trainer and continues to support the implementation of this literacy intervention. Her PhD research examined comprehension within reading acquisition processes, and the relationships between language, literacy and teaching interactions to support young children's engagement with text. Janet currently lectures in language and literacy education in undergraduate and postgraduate degree programs.

Debra Ives

Independent reviewer and analyst

Debra Ives brings many years of teaching experience to her current role as a researcher. She holds specialist qualifications in both the Teaching of English to Speakers of Other Languages (TESOL) and Early Childhood Education (ECE) and has been appointed to a range of classroom and specialist positions in primary and early childhood education settings across Brisbane and Melbourne. Debra is however best known for her research into ESL in Victorian primary schools and in her recently completed doctoral studies she investigated the differing ways in which generalist teachers perceive and respond to the educational needs of ESL learners in their classrooms.

Words 1–307 in order of frequency

I	for	ate	friend	really	old	funny	together	snake
the	dad	get	their	could	woke	book	walk	jumped
and	but	lived	put	shop	ball	bad	great	place
to	saw	am	gave	would	come	things	icecream	show
a	house	him	found	eat	ever	yesterday	loved	where
was	that	watched	from	fish	new	computer	magic	everyone
my	weekend	little	down	this	room	help	work	or
went	time	can	water	ran	nice	zoo	coming	shark
we	her	bought	party	first	scared	now	someone	something
on	go	brother	about	by	who	ride	team	asked
it	came	big	took	food	inside	castle	thing	OK
then	because	birthday	good	named	it's	toy	always	scary
he	up	them	other	baby	tree	cousins	boat	every
had	his	bed	see	cat	cake	look	red	walked
in	once	made	girl	outside	best	more	teacher	read
they	after	name	boy	away	fell	tried	its	world
with	fun	too	over	favourite	long	find	princess	monster
of	like	next	us	has	movie	four	shopping	slide
there	some	dog	your	family	soccer	I'm	until	thank
got	have	lots	off	lunch	how	happily	only	white
she	are	night	three	man	also	started	black	buy
said	out	not	dinner	shops	know	dragon	garden	dressed
played	going	friends	liked	football	last	much	still	fast
one	called	into	won	looked	sleep	rabbit	beautiful	head
is	all	an	morning	wanted	swimming	five	pool	walking
were	play	park	playing	bike	don't	turned	take	why
day	Sunday	will	want	no	just	another	well	blue *
so	upon	car	happy	lost	told	make	animals	dogs *
when	Saturday	our	what	TV	yes	cousin's	horse	footy *
home	did	do	as	fairy	around	breakfast	movies	here *
you	school	sister	love	cousin	lot	chips	names	killed *
at	two	be	if	stayed	today	door	bit	need *
me	very	people	again	Friday	beach	couldn't	race	playground *
mum	back	didn't	game	games	finished	present	sad	that's *
								watch *

* Words 299 to 307 occurred with equal frequency. This influenced the number of words in this list.

OXFORD
UNIVERSITY PRESS
AUSTRALIA & NEW ZEALAND

100 YEARS OF PUBLISHING
AUSTRALIA 1908-2008

www.oup.com.au