Journal of Phonetics & Audiology
ISSN: 2471-9455

Bretherton-Furness et al., J Phonet and Audiol
2016, 2:1
DOI: 10.4172/2471-9455.1000109

Commentary Open Access

Creating a Non-Word List to Match 226 of the Snodgrass Standardised Picture Set

Jess Bretherton-Furness*, David Ward and Douglas Saddy

School of Psychology and Clinical Language Sciences, University of Reading, Whiteknights Road, Reading, UK

***Corresponding author:** Jess Bretherton-Furness, School of Psychology and Clinical Language Sciences, University of Reading, Whiteknights Road, Reading, UK, Tel: +44 (0)118 378 6573; E-mail: J.bretherton-furness@pgr.reading.ac.uk

Received date: December 03, 2015; **Accepted date:** January 29, 2016; **Published date:** February 01, 2016

Copyright: ©2016 Bretherton-Furness J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Creating non-word lists is a necessary but time consuming exercise often needed when conducting behavioural language tasks such as lexical decisions or non-word reading. The following article describes the process whereby we created a list of 226 non-words matching 226 of the Snodgrass picture set [1]. In order to examine phoneme monitoring in fluent and non-fluent speakers we used the Snodgrass pictures created by Snodgrass and Vanderwart [1]. We also wished to look at phoneme monitoring in non-words so began creating a list of words that were matched to the Snodgrass pictures. The non-words created were matched on the following dimensions; number of syllables, stress pattern, number of phonemes, bigram count and presence and location of the target sound when relevant. These properties were chosen as they have been found to influence how easy or difficult it is to detect a target phoneme.

Rationale for creating a non-word list

The nature of non-words used in experimental work has been shown to be extremely important to the results of the study they're used for. For example, the more or less similar a non-word is to a real word effects the speed at which a lexical decision is made [2-5]. Gibbs and Van Orden [3] found that lexical decisions were fastest when the non-words used contained illegal letter strings – strings of letters that do not appear together in the language used e.g., /gtf/. Keuleers and Brysbaert [6], state that due to the impact non-words have on lexical decisions, they should only contain legal letter strings thus more closely approximating real words.

Phonotatic probability is the frequency with which different sound segments and segment sequences occur in the lexicon [7-11]. For example, /bl/ occurs commonly in English and is therefore thought to have a high phonotactic probability. It has been found that sensitivity to phonotactic probability develops in childhood and becomes increasingly sensitive as our lexicon grows [8,12-14]. Munson and Bable [15] suggested that this increase in sensitivity is reflective of our lexical representations becoming more segmental. As our lexicon expands, so too do the phonotactic possibilities and we become more sensitive to those segments which appear most often e.g., /bl/. Coady and Aslin [12] Storkel [8] and Zamuner, Gerken and Hammond [16] have found that phonotactic probability is reflected in the accuracy of speech in young children e.g. the lower the phonotactic probability the less accurate the speech. This finding, when applied to the two-step model of lexical access [17] can be explained in terms of the level of activation. When a speaker attempts to access a word in their lexicon this model proposes two steps, lemma retrieval and phonological retrieval. These two steps are not sequential and activation spreads throughout the retrieval network from semantic features to phonological features and back again. The most active phoneme units are then selected and positioned into the phonological frame. The model would suggest that those units with higher phonological probability have higher activation and are, therefore, more readily retrieved. For this reason it may be easier to detect /l/ when it is in a /bl/ combination rather than a /nl/ combination as /bl/ occurs more often in English than /nl/. As our list was created for a phoneme monitoring task controlling for the number of letter bigrams was especially important.

In Levelt et al., [18] model of speech production it is noted that we have the ability to monitor phonological code that is generated in the syllabification process which occurs before word production. Tasks such as phoneme monitoring can be used to test our ability to monitor phonological code which is what Schiller [19] did. Adult Dutch speakers were given a silent phoneme monitoring task in which the phoneme they had to monitor for occurred in the syllable initial and stress initial position and was compared to when it occurred in syllable initial but not stress initial position. It was found that phoneme monitoring occurs fastest when the phoneme occurs in the initial stress position. Dutch like English is a language in which the majority of multisyllabic words have their syllable stress on the initial syllable so results can be generalised to English. Coalson and Byrd [20] conducted a study asking participants to monitor for a phoneme in non-words. They found similar results to Schiller (2005) and also suggest that fluent adults monitor for phonemes more slowly in non-words as opposed to real words. It can be seen from this work that controlling for the position of the phoneme within the word and whether it occurs in the stressed syllable is important as it affects speed of monitoring.

Purpose of the list – current study

We created this non-word list as in our subsequent study we wished to examine phoneme monitoring in real and non-words in adult who are fluent vs. adults who are dysfluent. As we also wished to do this in a silent picture phoneme monitoring paradigm we chose to use the Snodgrass picture set [1]. Snodgrass and Vanderwart created this their set of 260 line drawings which they standardised on four variables; familiarity, image agreement, name agreement and visual complexity. These variables must be controlled for as they affect cognitive processing in pictorial and verbal form. More familiar items are more easily named as are words learnt at a younger age, those with higher name and image agreement, and less visual complexity, are also more easily named [21-23].

Generating the non-words

Initially we excluded some of the Snodgrass words e.g. those which are not regularly used in British English e.g. wrench (in English we

would use spanner) noun phrases were also excluded e.g., wine glass. We then transcribed each word orthographically and phonologically detailing position of primary stress, total number of syllables and the total number of phonemes. A letter bigram count was also calculated by hand. This count, taking account of phonological transcription, was vital as English orthographic transcription does not consistently agree with phonological transaction. Once we had all of this information we could begin creating our non-words.

In order to create the non-words we used two software programs. The first was the ARC Nonword Database [24]. This database was created so that researchers could access monosyllabic non-words or pseudo-homophones, chosen on the basis of a number of properties including; the number of letters, the neighbourhood size, summed frequency of neighbours, number of body neighbours, summed frequency of body neighbours, number of body friends, number of body enemies, number of onset neighbours, summed frequency of onset neighbours, number of phonological neighbours, summed frequency of onset neighbours, bigram frequency – type, bigram frequency – token (both position specific and position non-specific), trigram frequency – type, trigram frequency – token (both position specific and position non-specific) and the number of phonemes. Values for each of these can be set (upper and lower limits) and the fields you wish to have output for can also be selected. Non-words and pseudo-homophones can be chosen to be only orthographically existing onsets, be only orthographically existing bodies, only legal bigrams, monomorphemic only syllables, polymorphemic only syllables and morphologically ambiguous syllables. The ARC software, whilst extensive, could only be used to create non-words for all of the monosyllabic words in the Snodgrass set (121 words of the 226 total). Each word was chosen from a list of possible options given by the ARC database, when the target sound needed to be present non-words had to be selected that also had the target sound in the same position. It was not possible to ask the software to do this for us so added additional workload.

For the remaining 105 multisyllabic words we used the Wuggy software (Keuleers and Brysbaert, 2010) to create the non-words. Once again words were matched to real words in terms of, phoneme length, syllable length, presence or absence of the target sound, place in which the target sound occurred when it occurred and stress pattern. Wuggy is a multilingual pseudo-word generator designed to elicit non-words in Basque, Dutch, English, French, German, Serbian (Cyrillic and Latin), Spanish, and Vietnamese. This software was developed to expand upon what ARC offers as it can generate multisyllabic words. A word or non-word can be inputted and the algorithm can generate pseudo-words which are matched in sub-syllabic structure and transition frequencies. In the Wuggy software, after the language has been selected, it is possible to select whether real or pseudo-words are required. Output restrictions can then be applied including; match length of sub-syllabic segments, match letter length, match transition frequencies (concentric search) and match sub-syllabic segments e.g. 2 out of 3. There are also output options similar to ARC, including; syllables, lexicality, OLD 20, neighbours at edit distance, number of overlapping segments and deviation statistics. Each of the remaining 105 words were put into Wuggy and one of the options generated was chosen based upon whether it had the target sound (when applicable) in the correct location.

Once each non-word had been chosen and transcribed orthographically and phonologically a manual bigram count was taken. To ensure no bigrams were missed the total number of phonemes was calculated (980 phonemes in each list – words and non-words) following this the total number of possible bigrams was calculated (754 bigrams in each list – words and non-words). Bigram frequency data was calculated for real and non-words and a Wilcoxon signed rank test similar frequencies across the two word lists (z=-0.123, p=0.902). None of the non-words differed to the real words by more than 2 standard deviations (more than 5 bigrams) and the greatest difference was 6 occurrences of a bigram vs 1 occurrence of it. By ensuring that the lists are as similar as possible we have minimized the chance of any differences between performances on each list being down to factors other than the word/non-word distinction.

Outcome

The completed non-word list with corresponding Snodgrass words can be found in Table 1. The target phonemes that we used in the subsequent phoneme monitoring task are highlighted in bold (where applicable). It should be noted that whilst this list is matched and the bigram frequencies are such that there is no significant difference between the two lists, this is only the case when all 226 words are used. If exclusions are made in any work using them then a new bigram count must be taken to ensure that lists remain well matched.

S.NO.	Non-Word List	Non-Word List	S.NO.	Non-Word List	Non-Word List
1	əkɔ:di:ən	əfɑ:diən	115	bɑ:skɪt	bæskəl
2	eərəpleɪn	aɪrəʊtreɪt	116	bæt	bɒn
3	ælɪgeɪtə	ælaɪkætə	118	beə	ʃɔɪ
4	æŋkə	ælkɑ:	119	bed	pɪd
5	ænt	elt	120	bi:	θɑ:
6	æpəl	ʌpəl	121	bi:təl	si:təl
7	ɑ:m	i:m	122	Bel	vɪl
8	ærəʊ	eri:	123	belt	hent
9	ɑ:tɪtʃəʊk	æribɔ:k	124	baɪk	hi:k
10	æʃtreɪ	æʃtɑ:t	125	bɜ:d	beɪd
11	əspærəgəs	əspu:rərɒs	126	blaʊz	spɜ:tʃ
12	æks	keb	127	bʊk	dəʊk
13	bɔ:l	tʌl	128	bu:t	baʊn
14	bəlu:n	bəli:n	129	bɒtəl	bekəl
15	bənɑ:nə	ləmu:nə	130	baʊ	zeɪ
16	bɑ:n	vɔ:l	131	baʊl	hɒl
18	bærəl	sɑ:rəl	132	bɒks	sɪnt
19	bred	stɒd	133	i:gəl	elgə
20	bru:m	flæm	134	ɪə	u:
21	brʌʃ	fræʃ	135	elɪfənt	eməfens
22	bʌs	hes	136	envələʊp	enlədi:v
23	bʌtəflaɪ	bensəfi:	137	aɪ	əʊ

24	bʌtən	bɒθən	138	fens	pliːn
25	keɪk	səʊm	139	fɪŋgə	fænvə
26	kæməl	seməl	140	fɪʃ	teʃ
27	kændəl	sʌntəl	141	flæg	blɒf
28	kænən	mɑːnən	142	flaʊə	blaʊə
29	kæp	rɒp	143	fluːt	meɪnt
30	kɑː	zaʊ	144	flaɪ	klaɪ
31	kærət	ʃærɪt	145	fʊt	sɜːt
32	kæt	ket	146	fɔːk	gaɪk
33	kætəpɪlə	kætəbɜːgə	147	fɒks	swɪt
34	seləriː	bɪləni	148	frɒg	graːl
35	tʃeɪn	fep	149	dʒɪrɑːf	kɪræf
36	tʃeə	tʃeɪ	150	glɑːs	smɪʃ
37	tʃeriː	befiː	151	glɑːsɪz	dreɪsəs
38	tʃɪkɪn	tʃæzən	152	glʌv	stɒθ
39	tʃɪsəl	ʃæsəl	153	gaʊt	saʊn
40	tʃɜːtʃ	naːʃ	154	gərɪlə	kərəʊtʃə
41	sɪgɑː	pɪgaː	155	greɪps	drəʊks
42	sɪgəret	kɪpəraʊd	156	grɑːshɒpə	greslɜːpə
43	klɒk	stek	157	gɪtɑː	niːsɑː
44	klaʊd	smed	158	gʌn	sæn
45	klaʊn	bruːb	159	heə	ɔːn
46	kəʊt	hɜːk	160	hæmə	tæmə
47	kəʊm	dʒek	161	hænd	spæd
48	kɔːn	fiːn	162	hæŋə	tɑːnə
49	kaʊtʃ	rɜːp	163	hɑːp	tuːp
50	kaʊ	aʊn	164	hæt	sen
51	kraʊn	bræŋ	165	hɑːt	lɪtʃ
52	kʌp	lʌp	166	ʌnjən	ɪndən
53	dɪə	θaʊ	167	ɒrɪndʒ	ɒrɪntʃ
54	desk	lʌmf	168	ɒstrɪtʃ	ɒtrɪpt
55	dɒg	mʌp	169	aʊl	uːl
56	dɒl	næl	170	peɪntbrʌʃ	keɪntgrʌʃ
57	dɒŋkiː	mɒnveɪ	171	pitʃ	ʃʌf
58	dɔː	dɔɪ	172	pikɒk	duːʃel
59	dɔːnɒb	rɜːʃɒb	173	pinʌt	piːnɪl
60	dres	treɪdʒ	174	peə	nɜː
61	drʌm	slɒm	175	pen	hɪn

62	dʌk	kæz	176	pensəl	pɒnsəl
63	helɪkɒptə	hemɪteltə	177	peŋgwɪn	kengsuːn
64	hɔːs	laʊv	178	pepə	pɜːlə
65	haʊs	nʌs	179	piːænə	maɪəgaʊ
66	aɪən	eɪəm	180	pɪg	pæb
67	dʒækɪt	tʃɒket	181	paɪnæpəl	kaɪnæfəl
68	kæŋgəruː	sæŋgækiː	182	paɪp	feəp
69	ketəl	betəl	183	plaɪəz	klaɪəs
70	kiː	ɑːl	184	plʌg	lɒnt
71	kaɪt	jɒk	185	pəteɪtəʊ	pɪkeɪtə
72	naɪf	saːf	186	pʌmpkɪn	pɒmpkən
73	lædə	taʊdə	187	ræbɪt	pæbɪt
74	læmp	blɒp	188	ræku:n	sæku:n
75	liːf	wef	189	raɪnɒsərʊs	kraɪpɒkəbɑː
76	leg	wɒp	190	rɪŋ	vɜːn
77	lemən	tʃæmən	191	ruːlə	giːlə
78	lepəd	luːpəd	192	sɒlt	tɒlt
79	letɪs	kɜːrəs	193	sænwɪdʒ	sɑːknɪtʃ
80	laɪən	laiəl	194	sɔː	əʊl
81	lɪps	slʌd	195	sɪzəs	dʌzəs
82	lɒbstə	dɒbstə	196	skruː	bliːf
83	lɒk	lɔːk	197	skruːdraɪvə	tʃrɪbdraɪvə
84	mɪtən	fɪtən	198	sihɔːs	keəhɒs
85	mʌŋkiː	ræŋkiː	199	suːtkeɪs	suːlkæʃ
86	muːn	tʃæn	200	sʌn	kɒz
87	məʊtəbaɪk	kɑːtəpaɪk	201	swɒn	bræb
88	maʊntɪn	muːntɑːt	202	swetə	pliːtə
89	maʊs	gaʊs	203	swɪŋ	klaʊp
90	mʌʃruːm	kʌʃtuːm	204	teɪbəl	pæbəl
91	neɪl	maʊl	205	teləfəʊn	leməfeɪn
92	nekləs	gekləs	206	teləvɪʒən	feləsuːsən
93	niːdəl	widəl	207	θʌm	θɪm
94	naʊz	beɪm	208	taɪ	θuː
95	nʌt	gɪk	209	taɪgə	taɪdə
96	siːl	dʒɑːl	210	təʊstə	kuːstə
97	ʃiːp	ʃɜːp	211	taʊ	hɔɪ
98	ʃɜːt	saʊtʃ	212	təmɑːtəʊ	bəmɑːtuː
99	ʃuː	nɔɪ	213	tuːθbrʌʃ	kæŋbreʃ

100	skɜ:t	plaɪs	214	treɪn	preɪn
101	skʌŋk	trɪnk	215	tri:	trɔ:
102	sledʒ	gru:θ	216	trʌk	blæt
103	sneɪəl	flu:əl	217	trʌmpɪt	blempɪt
104	sneɪk	stæŋ	218	tɜ:təl	to:pəl
105	snəʊmæn	spaʊkæn	219	ʌmbrelə	ʌsfrɒlə
106	sɒk	fek	220	vɑ:s	bɑ:s
107	spaɪdə	brɪpə	221	vaɪəlɪn	baɪəʊmɪn
108	spu:n	trɔɪn	222	wɒtʃ	wæθ
109	skwɪrəl	skwɪrɪt	223	wɔ:təmelɒn	kɒtəmægən
110	stɑ:	tɒtʃ	224	wel	pel
111	stu:l	prɪl	225	wi:l	rɜ:l
112	staʊv	krəʊt	226	wɪndmɪl	wɪlmɪkt
113	strɔ:beri:	streɪbetʃi	227	wɪndəʊ	wændaʊ
114	-	-	228	zebrə	sɪbnə

Table 1: The completed non-word list with corresponding Snodgrass words.

References

1. Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn 6: 174-215.
2. Borowsky R, Masson ME (1996) Semantic ambiguity effects in word identification. J Exp Psychol Learn 22: 63-83.
3. Gibbs P, Van Orden GC (1998) Pathway selection's utility for control of word recognition. J Exp Psychol Hum Percept Perform 24: 1162-1187.
4. Gerhand S, Barry C (1999) Age-of-acquisition and frequency effects in speeded word naming. Cognition 73: B27-B36.
5. Ghyselinck M, Lewis MB, Brysbaert M (2004) Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. Acta Psychologica 115: 43-67.
6. Keuleers E, Brysbaert M (2010) Wuggy: A Multilingual pseudoword denerator. Behaviour Research Methods 42: 627-633.
7. Jusczyk PW, Luce PA (1994) Infants′ sensitivity to phonotactic patterns in the native language. J Mem Lang 33: 630-645.
8. Storkel HL (2001) Learning New Words Phonotactic Probability in Language Development. J Speech Lang Hear Res 44: 1321-1337.
9. Storkel HL (2003) Learning New Words II: Phonotactic Probability in Verb Learning. J Speech Lang Hear Res 46: 1312-1323.
10. Vitevitch MS, Armbruster J, Chu S (2004) Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. J Exp Psychol Learn Mem Cogn 30: 514-529.
11. Vitevitch MS (2002) The influence of phonological similarity neighborhoods on speech production. J Exp Psychol Learn Mem Cogn 28: 735-747.
12. Coady JA, Aslin RN (2004) Young children's sensitivity to probabilistic phonotactics in the developing lexicon. J Exp Child Psychology 89: 183-213.
13. Edwards J, Munson B, Beckman M (2004) Nonword Repetition in Children with Phonological Disorders. Poster presentation at the Symposium for Research on Child Language Disorders, Madison, WI.
14. Munson B, Kurtz BA, Windsor J (2005) The Influence of Vocabulary Size, Phonotactic Probability, and Wordlikeness on Nonword Repetitions of Children with and without Language Impairments. J Speech Lang Hear Res 48: 1033-1047.
15. Munson B, Babel ME (2005) The sequential cueing effect in children's speech production. Applied Psycholinguistics 26: 157-174.
16. Zamuner TS, Gerken L, Hammond M (2004) Phonotactic probabilities in young children's speech production. J Child Lang 31: 515-536.
17. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA (1997) Lexical access in aphasic and nonaphasic speakers. Psychol rev 104: 801-838.
18. Levelt WJM, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. Behav Brain Sci 22: 1-38.
19. Schiller NO (2005) Verbal self-monitoring. In A. Cutler (Edn.) Twenty-first century psycholinguistics: Four cornerstones, Mahwah NJ: Erlbaum pp. 245-261.
20. Coalson GA, Byrd CT (2015) Metrical Encoding in Adults Who Do and Do Not Stutter. J Speech Lang Hear Res 58: 601-621.
21. Ellis AW, Morrison CM (1998) Real age-of-acquisition effects in lexical retrieval. J Exp Psych: Learning Memory and Cognition 24: 515.
22. Funnell E, Sheridan J (1992) Categories of knowledge? Unfamiliar aspects of living and nonliving things. Cognitive Neuropsychology 9: 135-153.
23. Gilhooly K, Gilhooly M (1979) Age-of-acquisition effects in lexical and episodic memory tasks. Memory & Cognition 7: 214-223.
24. Rastle K, Harrington J, Coltheart M (2002) 358,534 nonwords: The ARC Nonword Database. The Quarterly Journal of Experimental Psychology Section A 55: 1339-1362.