Towards a standardized, fine-grained manual annotation protocol for verbal fluency data

Gabriel Frazer-McKee[1,2], Joël Macoir[1,2], Lydia Gagnon[1] & Pascale Tremblay[1,2]

[1]CERVO Brain Research Center, Quebec City, Canada
[2]Faculty of Medicine, Department of Rehabilitation, University Laval, Canada

Abstract

We propose a new method for annotating verbal fluency data, which allows the reliable detection of the age-related decline of lexical access capacity. The main innovation is that annotators should inferentially assess the intention of the speaker when producing a word form during a verbal fluency test. Our method correlates probable speaker intentions such as "intended as a valid answer" or "intended as a meta-comment" with linguistic features such as word intensity (e.g. reduced intensity suggests private speech) and syntactic integration. The annotation scheme can be implemented with high reliability, and minimal linguistic training. When fluency data are annotated using this scheme, a relation between fluency and age emerges; this is in contrast to a strict implementation of the traditional method of annotating verbal fluency data, which has no way of dealing with score-confounding phenomena because it force-groups all verbal fluency productions –regardless of speaker intention— into one of three taxonomic groups (i.e. valid answers, perseverations, and intrusions). The traditional lack of fine-grained annotation units is especially problematic when analyzing the qualitatively distinct fluency data of older participants and may cause studies to miss the relation between lexical access capacity and age.

1 Introduction

Verbal fluency tasks (also known as verbal fluency tests; VFTs) are hybrid neuropsychological tasks (Shao et al., 2014) that assess both lexical access capacity –the capacity to retrieve lexical units from the mental lexicon (Indefrey & Levelt, 2004)— and executive function. In VFTs, participants are required to provide as many (non-repeated) words as possible that correspond to the experimental criteria within a time-limit (typically, 60 seconds). These criteria are either semantic (e.g. name as many animals as possible) or phonemic (e.g. list words that begin with the letter *P*). VFTs are amongst the most widely used language tasks in both research and clinical settings owing to their sensitivity to a wide variety of psycho-neurobiological phenomena –dementia (e.g. Troyer et al., 1997), bilingualism (e.g. Sandoval et al., 2010), and aging (e.g. Gordon et al., 2017), to name but these. Performance on these tasks has been used to characterize the fluency capacity of various groups, the classic performance measures being the number of correct answers, the number of intrusions (words that do not meet the experimental criteria), and the number of perseverations (correct answers that have been repeated) (e.g. Strauss et al., 2006).

While the traditional three-way taxonomy (i.e. correct answers, perseverations, intrusions) is easy to use and is widespread in the fluency literature (e.g. Ledoux et al., 2014; Strauss et al., 2006; Troyer et al., 1997), it has a serious shortcoming: it can only handle clean data. Consider, for instance, the following illustrative example of a series of verbal productions drawn from the T fluency condition. Font size is used to suggest speech intensity (volume), ellipsis is used to represent pauses, colons denote an elongated speech sound, dashes are used to denote temporal proximity, and forward slashes represent truncation; the data of interest is underlined:

1. Participant F027: *Tama/—tamarin!*
   [Tama (type of drum) tamarind! → **false start**]

2. Participant F038: *tro/ Euh trottoir je l'ai dit*
   [side—! uhm, I already said *sidewalk* → **intentional repetition found in a comment**]

3. Participant F037: *trois! ta ...: ta: ... tapis!*
   [three! your … your … rug → **syllable play**]

The above illustrates some of the interesting, but potentially score-confounding phenomena found in fluency data: presumably attempts to connect contextually meaningless syllables (e.g. *ta: ta:*) with words stored in the mental lexicon (example 3 above), very much intentional repetitions (example 2 above), and unintentional false starts that happen to correspond to actual words (example 1 above).

Existing fluency protocols completely gloss over these types of phenomena (e.g. Strauss et al., 2006), but in our own fluency corpus we have observed that such phenomena are not unusual amongst cognitively healthy participants, especially older ones (see subsection 3 for some descriptive statistics). Importantly, it is well-documented that older adults experience not only a decline in lexical access but also have different narrative styles compared to younger adults, particularly when the task is difficult (Mortensen et al., 2006). It is therefore possible that older adults' verbal fluency data differs qualitatively compared to that of younger adults; to our knowledge, no verbal fluency study has investigated this issue to date. The open question is whether there are sufficient instances of these score-confounding phenomena in the fluency data to impact performance scores (and thus study findings) in one or more fluency conditions, and more specifically whether the performance measures in these conditions vary with age when score-confounding phenomena are controlled for. What is needed to investigate this research question is to separate the data of interest from score-confounding phenomena using a standardized, fine-grained taxonomy that allows fluency annotators to make replicable, theoretically-justified categorization choices based on linguistic features that correlate with specific, contextually-plausible communicative intentions. In this paper, we briefly describe such a taxonomy and report promising preliminary findings bearing on a) the reliability of our protocol's transcription-annotation process, and b) the effect this taxonomy has on two classic measures of fluency performance in younger and older adults compared to a strict implementation of the traditional annotation scheme.

2 Creation of the fluency corpus

To date, 38 depression-free (GDS-15≥5; Sheikh & Yesavage, 1986), cognitively-normal (MoCA≥26; Nasreddine et al., 2005), right-handed (Oldfield≥14; Oldfield, 1971) native speakers of French (age=62.3 years; SD=17.08; range=28-88; females=21; schooling=14.65 years; SD=2.17) have been audio-recorded during the execution of 4 verbal fluency tasks (letters: T-N-P; category: animals). Participants are classified into one of two age groups: younger (n=14; M=40.48 years; SD=8.76; range=22-54) and older (n=24; M=73.44 years; SD=6.71; range=65-88). Participants were recorded in a double-walled soundproof room (Génie Audio. Inc, Canada) using a Shure headset microphone (Microflex Beta 53) connected to a Quartet USB audio interface (Apogee Electronics, Santa Monica, CA 90404, USA) that fed into an iMac computer. This permitted the capture of very low-intensity phenomena (e.g. whispering). The recordings were made using Sound Studio 4 (Felt Tip Inc., NYC, USA) at a sampling rate of 48 kHz with 24 bits of quantization. Per traditional procedure (e.g. Strauss et al., 2006), participants were instructed to produce as many words as possible that corresponded to the experimenter-provided criteria within 60 seconds, avoiding repetitions, proper names, and words that merely have different suffixes (e.g. *pitch/pitched* or *cat/cats*). The fluency tasks were administered as a part of a battery of tests within a large on-going project that was approved (#1733-2018) by the neurosciences and health ethics review board of the *Centre intégré universitaire de la santé et des services sociaux de la Capitale Nationale* (translation: Integrated Health and Social Services Centre of the National Capital) in Quebec City, Canada. The raw fluency audio recordings were transcribed and annotated in Praat

(Boersma & van Heuven, 2001) using the extended version of the transcription-annotation scheme described below (see Appendix B).

3 Towards a finer-grained annotation scheme: The linguistic correlates of intent

Our expanded fluency taxonomy rests on two fundamental assumptions. Firstly, a fluency taxonomy should be comprised of categories that reflect speaker communicative intentions. Simply put, the issue is not whether the *fluency annotators* can associate –using all knowledge at their disposal— a given phonetic sequence within the data with conceptual content and thereby categorize the sequence using some rule-scheme. The issue, fundamentally, is whether one can reasonably infer that the *speaker* associated the sequence with conceptual content at the moment of verbalization, and, what is more, whether they produced the sequence to *satisfy* the experimental criteria (intentionality).

Our second fundamental assumption is that speaker intent has linguistic correlates (cf. Sperber & Wilson, 2002). Using a corpus-driven approach (Biber, 2009), we associated probable pragmatic intentions (e.g. intent to satisfy the experimental criteria, etc.) with syntactic and phonetic features. The taxonomy's key categories are briefly described below; a full list can be found in Appendix B. Percentage frequency of occurrence of each of the annotation units is featured in parentheses in the manner that follows: relative frequency in the whole corpus, relative frequency in the fluency data of adults aged 28-54, and relative frequency in the fluency data of adults aged 65+. Features that served to make annotation decisions are also included in each description.

1. **Correct answers** (56.10% / 28-54: 59.06% / 65+: 54.51%): verbal productions that a) meet the category-criteria, and b) were intended by the speaker to satisfy the experimental criteria (i.e. to be answers); they are fully pronounced words that are typically the loudest verbal productions in the data (per visual inspection of the spectrograms of 100 randomly selected datapoints from our fluency corpus); they tend to cluster with other correct answers and/or vocalics.
2. **Perseverations** (2.21% / 28-54: 1.19% / 65+: 2.82%): repetition of a previous answer (often with intervening linguistic material between the two instances); participants are sometimes aware of these mistakes (cf. Day, 1979). *Conscious* perseverations are typically followed by a) a vocalic that indicates error awareness or b) a meta-comment such as "*said it*". Conscious perseverations can also be truncated (e.g. *side—uhm [sidewalk]*). Otherwise, perseverations usually have the phonetic/syntactic characteristics of correct answers.
3. **Vocalics** (23.43% / 28-54: 27.22% / 65+: 20.80%): paralinguistic verbal productions, such as sighs, hesitations, frustration noises, etc. (Burgoon et al., 2016); they often occur between (clusters of) answers (where they indicate active search) or follow truncated phonemic sequences (where they serve to evaluate the verbal production).
4. **Self-talk** (2.31% / 28-54: 1.26% / 65+: 2.93%): (repeated) (correct) answers that are instances of "thinking-out-loud" during a cognitive task (cf. Duncan & Cheyne, 2001) rather than perseverations per se, as evidenced by a) (dramatically) lowered voice intensity –a feature associated with the notion of privacy (Cirillo, 2004)— or b) syntactic integration with a temporally proximate meta-comment (e.g. *uhm, I said sidewalk*);
5. **Meta-comments** (6.89% / 28-54: 5.97% / 65+: 7.42%): comments on the participant's own performance or on something the participant just said (e.g. *aunt –person; ant –insect*) or a remark/question on the experimental task itself (McDowd et al., 2011; Roberts & Le Dorze, 1997); often delivered at a comparatively rapid rate of speech and considerably lower intensity than correct answers, per visual inspection of the spectrograms of 50 randomly selected meta-comments from our fluency corpus.
6. **Syllable play** (5.02% / 28-54: 1.46% / 65+: 7.02%): a syllable in the participant's language that could be considered as a word (i.e. conventional form-meaning pairing) but that is either a) a false start (e.g. *tama—tamarind*) or that b) the participant has likely failed to associate with lexical content, as evidenced by following meta-comments or acoustic characteristics such as vowel elongation and reduced speech intensity (e.g. *pa: pa: Ah, come on!*). Instances of syllable play tend to be clustered together or to occur shortly before a correct answer/perseveration.

4 Testing the new annotation scheme

In the following, we assess our fluency annotation scheme's reliability as well as its effect on two classic measures of fluency performance. We also show that our fluency protocol is sensitive to age-related decline of lexical access, and that this decline emerges largely because of the way in which syllable play and self-talk are handled by our inferential annotation scheme.

4.1 Measures of agreement

17.5% of the transcriptions (n=7 participants; 823 unique datapoints) and 47.5% of the annotations (n=19 participants; 2062 datapoints) have been independently analyzed by two annotators. Raw agreement for both the transcriptions and annotations is excellent, averaging respectively 95.26% and 99.38% between the 4 fluency conditions (see Tables 3 and 4 in Appendix A for condition-specific details). The overall intra-class correlation coefficient (ICC) for the annotations of the fluency corpus is also excellent (0,987; see Table 4 in Appendix A for condition-specific details), per well-known magnitude interpretations of ICCs (e.g. Koo & Mae, 2016). Our annotation protocol's ICC is very comparable to the ICCs of the traditional taxonomy (e.g. ≈0.98 in Passos et al., 2011). Thus, taking considerably more phenomena into consideration does not adversely affect the accuracy or the reliability of the transcription-annotation process.

4.2 Comparison of the two annotation protocols

The true test of our taxonomy is whether it offers concrete empirical advantages –for instance, whether it reveals different associations between age groups and performance measures compared to an annotation scheme that has no way of handling score-confounding phenomena.

A linear mixed model with participants as random intercepts was used to evaluate the interaction between age (inter-subject categorical variable), annotation system (intra-subject variable), and number of correct answers (dependent variable) in each of the fluency conditions (N, P, T and Animals). The traditional system was found to yield more correct answers than our protocol in the T ($p < 0.001$) and N ($p > 0.01$) fluency conditions (Table 1A).

	A. Number of correct answers	**B. Number of perseverations**
	(Linear Mixed Model)	(Mann-Whitney U test)
N	$p < 0.000$	$p < 0.048$
P	assumption violation	$p < 0.033$
T	$p < 0.01$	$p < 0.018$
Animals	assumption violation	$p < 0.043$

Table 1. Comparison of two fluency protocols along two classic performance parameters: number of correct answers and number of perseverations (α=0.05)

What is more, a statistically significant interaction was found between age and type of protocol in the T ($p < 0.05$) and N ($p < 0.010$) conditions: while the traditional taxonomy suggests that fluency performance is stable in these conditions with age, our inferential system suggests that phonemic fluency performance declines slightly with age.

Since a) the perseveration data were not distributed normally and b) were relatively infrequent to begin with, it was not possible to use a linear mixed model to assess their relationship with taxonomy-type and participant age. We therefore performed a Mann-Whitney U test to determine whether the two protocols (independent variable) yield significantly different perseveration scores (dependent variable) in each of the fluency conditions. The traditional taxonomy was found to produce significantly higher perseveration scores in all of the fluency conditions (see Table 1B).

As was predicted, the protocol-x-age interaction effect was found to largely be due to the prevalence of self-talk and syllable play in the two age groups and how these data are respectively handled by the two protocols. Significant differences were found between younger and older participants' use of syllable play and self-talk, per Table 2 below:

	Syllable play		Self-talk	
	28-54 (n=14)	65+ (n=24)	28-54 (n=14)	65+ (n=24)
N	14.29%	62.50%	14.29%	50.00%
P	21.42%	50.00%	21.42%	16.67%
T	35.71%	75.00%	21.42%	58.33%
Animals	0.00%	0.00%	14.29%	48.00%

Table 2. Percentage of participants who produced at least one instance of syllable play/self-talk according to age-group (28-54 vs. 65+)

These two categories account for only just over 7% of the corpus (see subsection 3 of this article), but are not evenly distributed according to age-group. Compared to the younger participants, older participants were between 1.9x to over 4x more likely to produce at least one instance of self-talk or syllable play in all but one fluency condition (i.e. self-talk in the P fluency condition).

5 Discussion and future work

The traditional fluency taxonomy has been used for decades but it offers no formally codified means to avoid the force-grouping of data that have fundamentally different cognitive causes (e.g. correct answers and syllable play), and whose use may, crucially, differ across the adult lifespan. Given that fluency scores are typically based on a mere 60 seconds of verbal production and that differences in fluency performance between groups are often modest to begin with (e.g. Macoir et al., 2019), the non-systematic treatment of score-confounding phenomena has the potential to alter individual performance scores (and thereby study results). The accuracy of the annotation process is thus not a trivial matter, particularly for studies with small effect sizes and/or studies with small sample sizes (where annotation choices carry comparatively greater statistical weight). If fluency studies are to be maximally consistent and their results maximally comparable, the handling of fluency data cannot fall under the scope of "researcher degrees of freedom" –undeclared "flexibility in data collection, analysis, and reporting" (Simmons et al., 2011, p. 1359). A detailed, publicly accessible annotation protocol –one that weighs the linguistic evidence that bears upon the participant-speaker's communicative intentions— is required to separate the fluency data into finer-grained categories in a non-arbitrary manner so that cognitively similar phenomena of interest may be identified and isolated for the purpose of statistical analysis. The extended fluency taxonomy briefly described and tested here constitutes a promising first step towards such a standardized transcription-annotation fluency protocol. Preliminary results suggest that a) the protocol is highly implementable, as evidenced by excellent agreement and reliability scores; b) it requires minimal linguistic training (as evidenced by its successful use by LG, a first year undergraduate Linguistics student); and c) most importantly, it yields non-inflated fluency performance scores compared to a purely mechanistic implementation of the traditional fluency taxonomy. Preliminary findings suggest that the adoption of such a protocol is particularly germane for the comparative analysis of the verbal fluency data of adult participants of different ages. Based on our corpus here, we suggest that older adults tend to produce qualitatively distinct fluency data compared to younger adults (likely as a compensatory strategy for declining lexical access ability), hence why the two annotation schemes produce significantly dissimilar performance measures for this group).

With a new, finer-grained annotation scheme to structure the data also come new, exciting research questions. Does the use of fillers and self-talk in fluency data vary along sociodemographic and/or cognitive parameters? What proportion of fluency errors are cognitively normal participants aware of? How effective is syllable play as a word identification strategy and is its effectiveness moderated by age? Do older participants who use lexical access strategies such as self-talk and syllable play achieve lesser or greater performance scores according to our fluency protocol? Perhaps most importantly of all, might the controversy regarding the maintenance or decline of phonemic fluency performance over the lifespan (cf. Gordon et al., 2017) be partly attributable to inconsistent annotation procedures between fluency studies? We intend to soon investigate these and other questions using a considerably expanded fluency corpus (120-150 participants).

Acknowledgements

Authors’ contributions: Gabriel Frazer-McKee (data collection and analysis, study conceptualization, article drafting and revision); Joël Macoir (study conceptualization, article revision); Lydia Gagnon (data collection and analysis, article revision); Pascale Tremblay (funding acquisition, project supervision, study conceptualization, article revision).

The authors thank the following for their contributions: Valérie Brisson (statistical advice), Elisabeth Maillard (data collection), Alison Arseneault (data collection), Sophie Simard (data collection), Catherine Savard (data collection), Julia Picard (data collection), and all of the study’s participants. The authors would also like to sincerely thank the two anonymous evaluators for their insightful and encouraging comments on the submitted version of this article.

This work was supported by a grant to Pascale Tremblay from the Fonds de recherche Nature et Technologies du Québec (2019-PR-254714), and a Globalink research internship from MITACS. Pascale Tremblay holds a Career award from the *Fonds de la Recherche en Santé du Québec* (FRQ-S, #35016).

References

Biber, D. (2009). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 193–224). Oxford University Press.

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9–10), 341–347.

Burgoon, J. K., Guerrero, L. K., & Manusov, V. (2016). *Nonverbal Communication*. Routledge.

Cirillo, J. (2004). Communication by unvoiced speech: The role of whispering. *Anais Da Academia Brasileira de Ciências*, *76*(2), 413–423.

Day, R. S. (1979). Verbal fluency and the language-bound effect. In C. J. Fillmore, D. Kempler, & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 57–84). Academic Press.

Duncan, R. M., & Cheyne, J. A. (2001). Private speech in young adults: Task difficulty, self-regulation, and psychological predication. *Cognitive Development*, *16*(4), 889–906.

Gordon, J. K., Young, M., & Garcia, C. (2017). Why do older adults have difficulty with semantic fluency? *Aging, Neuropsychology, and Cognition*, *25*(6), 1–26.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Koo, T. K., & Mae, Y. L. (2016). A guideline of selecting and reporting Intraclass Correlation Coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Ledoux, K., Vannorsdall, T. D., Pickett, E. J., Bosley, L. V., Gordon, B., & Schretlen, D. J. (2014). Capturing additional information about the organization of entries in the lexicon from verbal fluency productions. *Journal of Clinical and Experimental Neuropsychology*, *36*(2), 205–220.

Macoir, J., Lafay, A., & Hudon, C. (2019). Reduced lexical access to verbs in individuals with Subjective Cognitive Decline. *American Journal of Alzheimer’s Disease & Other Dementias*, *34*(1), 1–15.

McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., & Kemper, S. (2011). Understanding verbal fluency in healthy aging, Alzheimer’s disease, and Parkinson’s disease. *Neuropsychology*, *25*(2), 210–225.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes*, *21*(1–3), 238–290.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113.

Passos, V. M. de A., Giatti, L., Barreto, S. M., Figuereido, R. C., Caramelli, P., Benseñor, I., & Trinidade, A. A. M. da. (2011). Passos, V. M. de A., Giatti, L., Barreto, S. M., Figueiredo, R. C., Caramelli, P., Benseñor, I., … Trindade, A. A. M. da. (2011). Verbal fluency tests reliability in a Brazilian multicentric study, ELSA-Brasil. , 69(5), 814–816. Doi:10.1590/s0004-282x2011000600017. *Arquivos de Neuro-Psiquiatria*, *69*(5), 814–816.

Roberts, P. M., & Le Dorze, G. (1997). Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain & Language*, *59*(3), 412–449.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*(772), 1–10.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*, *5*(1–2), 165–173.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11).

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language*, *17*(1), 3–33.

Strauss, E., Sherman, E. S., & Spreen, O. (2006). Verbal fluency. In *A compendium of neuropsychological tests: Administration, norms, and commentary* (pp. 499–526). Oxford University Press.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146.

Appendix A. Detailed agreement measure tables.

	Agreements	Disagreements	Raw agreement
N	164	14	92.13%
P	184	4	97.87%
T	175	11	94.09%
Animals	261	14	94.91%
Total	784	39	95.26%

Table 3. Inter-rater agreement for transcription decisions based on a subsample of the fluency corpus' participants (n=7)

	Agreements	Disagreements	Raw agreement	*ICC
N	424	4	99.07%	0.974
P	472	4	99.15%	0.991
T	506	2	99.61%	0.997
Animals	647	3	99.56%	0.983
Total	2049	13	99.37%	0.987

Table 4. Inter-rater agreement and reliability for annotation decisions based on a subsample of the fluency corpus' participants (n=19)

*ICC=Intra-class Correlation Coefficient (average measure)

Appendix B. The current annotation scheme (in alphabetic order).

* diacritic appended to errors to indicate participant error awareness (e.g. *INTRU); **CONTINU** Continuous perseveration (e.g. *Bee! ...Bee*). **CORR** Correction (e.g. repetition of a word with clearer or with more prestigious pronunciation). **F** false start (e.g. *p— pr—*). **HMPHN** Homophone (a type of correct answer). **INTRU** Intrusion. **META** Meta-comment. **NONCE** Nonce word. **PHON** Phonological error (e.g. cat + dog → *cog!*). **PRP** Proper name (type of intrusion). **RECUR** Recurrent perseveration (e.g. *cow, ox, cow*). **SELF** Self-talk. **SYLLAB** Syllable (play). **VAL** Correct answer; **VOC** Vocalic.