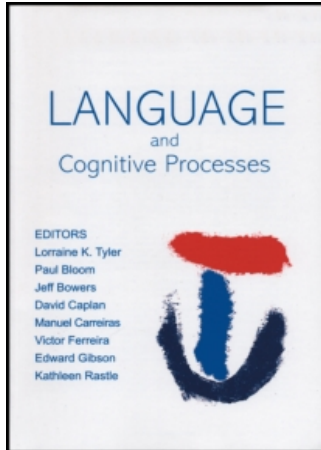


This article was downloaded by:[Max Planck Inst & Research Groups Consortium]  
On: 4 December 2007  
Access Details: [subscription number 771335669]  
Publisher: Psychology Press  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title~content=t713683153>

### Picture naming and word frequency: Comments on Alario, Costa and Caramazza, Language and Cognitive Processes , 17(3), 299-319

Willem J. M. Levelt

Online Publication Date: 01 December 2002

To cite this Article: Levelt, Willem J. M. (2002) 'Picture naming and word frequency: Comments on Alario, Costa and Caramazza, Language and Cognitive Processes , 17(3), 299-319', Language and Cognitive Processes, 17:6, 663 - 671

To link to this article: DOI: 10.1080/01690960143000443

URL: <http://dx.doi.org/10.1080/01690960143000443>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

**Picture naming and word frequency:  
Comments on Alario, Costa and Caramazza,**  
*Language and Cognitive Processes*, 17(3), 299–319

Willem J. M. Levelt

*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

This commentary on Alario et al. (2002) addresses two issues: (1) Different from what the authors suggest, there are no theories of production claiming the phonological word to be the upper bound of advance planning before the onset of articulation; (2) Their picture naming study of word frequency effects on speech onset is inconclusive by lack of a crucial control, viz., of object recognition latency. This is a perennial problem in picture naming studies of word frequency and age of acquisition effects.

Alario, Costa and Caramazza (2002) report two experiments on the production of determiner–adjective–noun phrases, such as “the blue kite”. These utterances were responses to pictures of coloured objects, a blue kite in the example. Their dependent measure was speech onset latency, the interval from picture onset to speech onset. As independent measures they orthogonally varied the word frequencies of the colour adjectives and of the nouns. In both experiments they obtained additive effects of these independent variables. When the adjective or the noun in an utterance was low frequency, speech onset latency was longer than when it was high frequency. From the fact that not only the first phonological word, but also the second one produced a frequency effect they concluded: “the results are problematic for models which assume that frequency affects the level of phonological encoding *and* that the scope of phonological encoding is limited to the first phonological word” (p. 315). In the following I will first argue that none of the cited authors assume that phonological encoding is limited to the first phonological word. I will then turn to the experiments

---

Requests for reprints should be addressed to Willem J. M. Levelt, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Email: pim@mpi.nl  
I am grateful to Ardi Roelofs for a careful reading of the manuscript.

and show that they are inconclusive by lack of an essential control, a serious defect shared with other work in this area.

## LOWER AND UPPER LIMITS OF PHONOLOGICAL ENCODING

When a speaker produces an utterance such as “the blue kite” in response to a picture, various levels of processing are involved. The speaker must recognise the picture, in particular the object represented and its colour. The speaker must select the relevant lexical items, the two content words and the determiner with their syntactic properties; the items’ syntax determines their ultimate order in the utterance. The speaker must retrieve the word forms, or phonological codes, of all three selected items and syllabify the utterance appropriately (/ðə.blu:.kaɪt/). The incremental syllabification guides the speaker’s computation of the successive articulatory motor gestures, whose execution by the articulatory system will produce the overt utterance. Each of these levels of processing involves its own characteristic processing units, or at least this is a long-standing topic of research. At the level of phonological encoding (including the retrieval of phonological codes and syllabification) a much discussed unit is the phonological or prosodic word ( $\omega$ ). Any domain of syllabification forms a phonological or prosodic word (Booij, 1995).<sup>1</sup> The domain of syllabification can be larger or smaller than a lexical word (the word in its citation form). For instance, in the utterance “the students understand it”, “understand it” forms a phonological word; it syllabifies as *un.der.stan.dit*, the last syllable straddling the boundary between two lexical words, “understand” and “it”. The utterance “the blue kite” will be produced as two phonological words, “the.blue” and “kite”. Small, unaccented words tend to “cliticise” to a neighbouring content word. Alario et al. (2002) share this conception of the phonological word (“a content word plus the function words that ‘attach’ to it”).

The domain of syllabification can also be smaller than the lexical word. Following Booij (1995), this is generally the case for compounds in Dutch. Each morpheme or root in a compound forms its own domain of syllabification, hence each such constituent is a phonological or prosodic word ( $\omega$ ) of its own. The same holds for most compounds in English. Take “landowner” as an example. If this compound would be a single phonological word, it would syllabify as *lan.dow.ner*, just like

---

<sup>1</sup> Nespor and Vogel (1986) distinguish the phonological word from the “clitic group”. Others, such as Booij (1995), argue against making this finer distinction. Here we follow Booij in that the syllables in a domain of syllabification form one prosodic (or phonological) word. In that respect, the domain of syllabification defines the phonological word.

“landau”—*lan.dau*. But it does not; “land” and “owner” are kept apart as domains of syllabification. Generally, the prosodic make-up of two-root compounds in Dutch or English is of the type  $\omega_1\omega_2$ .

It is a dominant (though not yet universally shared) assumption in production theory that the speed of phonological encoding is word frequency dependent. In particular, accessing a lexical word’s phonological code appears to be faster for high-frequency words than for low-frequency words (Jescheniak & Levelt, 1994; Levelt, Roelofs, & Meyer, 1999). Alario et al. (2002) mention other levels of processing that may be frequency-sensitive, but here I will limit myself to the phonological account of word frequency. As far as this account is concerned, the authors’ argument is straightforward: If utterance onset in the picture naming experiments is affected by noun frequency, then the speaker must have accessed the code for the second phonological word (“kite” in the example) before speech onset. And that is indeed the experimental result. This, the authors argue, is problematic for models that assume that “phonological planning units do not comprise more than one phonological word” (p. 312), the phonological word as an upper bound on advance phonological planning. That is correct, but alas, there are no such theories around. The authors refer to Levelt (1989). There one can read that ‘phonological phrases are important units of phonological encoding’ (p. 420). I argue in particular that phonological phrases consist of one or more phonological words and that such phrases are stored for word by word delivery. They also refer to Levelt et al. (1999). That article presents a theory of (phonological) word production, not of phrase production. The issue of how many phonological words are planned before speech onset is not discussed. The paper does discuss whether the phonological word is a lower bound, i.e., whether there will be no onset of articulation before a phonological word’s syllabification is completed. It concludes that “The evidence, however, is so far insufficient to make this a strong claim” (p. 33). Alario et al. further refer to Meyer (1996). In that paper one can read that “speakers probably use different planning units in different situations” (p. 480) and “It remains to be determined which of these processes [of phonological encoding—WL] are sequential within and between words” (p. 492). Her experimental results show the second phonological code to be active before speech onset in a subset of cases. Hence, there is no claim that the phonological word forms the upper bound. They refer to Schriefers and Teruel (1999), but there one can read: “Does this imply that the phonological word represents a fixed upper limit for the size of the phonological advance planning unit? The results of Sternberg et al. (1988) show that this is not the case” (p. 46). And on p. 47: “Taken together, the present results do not provide evidence for a phonological planning unit of some fixed size”. Finally, Alario et al. (2002) repeatedly refer to Wheeldon & Lahiri (1997) as representing the view that

the upper bound of advance phonological planning is a single phonological word. That is particularly surprising because these authors provided experimental evidence for the claim that the number of phonological words in an utterance is a determinant of utterance onset latency. They show that, everything else controlled, subjects were slower in producing a 3- $\omega$  utterance than a 2- $\omega$  one, and slower in producing a 4- $\omega$  utterance than a 3- $\omega$  one. Apparently subjects had prepared up to three, respectively four phonological words before utterance onset. Wheeldon and Lahiri used these (and other) data to argue for the significance of  $\omega$  as a planning unit, but they precisely argue against  $\omega$  to be the upper bound on advance planning. It should be added, though, that these results were obtained in a prepared speech paradigm, roughly following Sternberg et al. (1978). Here subjects uttered the planned utterance after a “go” signal. When the production was immediate, without delay, the complexity of the first phonological word, but not of the second one, affected the speech onset latencies. This shows that in certain cases encoding just the first phonological word of an utterance *can* be sufficient for the initiation of speech. The phonological word *can* be the minimal encoding unit before articulation is initiated, as argued by Levelt (1989, p. 421). In that case, the “latency to produce a sentence should be a function of the time required to construct the first phonological word rather than a function of the total number of phonological words it contains” (pp. 374–375). Taken together, however, the Wheeldon and Lahiri results show that the size of a speaker’s advance planning is variable, dependent on the given task. They conclude, correspondingly, “that when it is possible to do so, speakers preferentially initiate articulation following the phonological encoding of the initial phonological word of an utterance” (p. 377). Hence, a preference, not an upper bound. The “prosodic encoding of these sentences [the ones in the non-delayed production experiment—WL] required minimal processing with regards to larger prosodic structures and intonation. It is therefore possible that with longer and more complex sentences effects of whole sentence complexity may be observed in on-line sentence production latency” (p. 377). In short, Alario et al. (2001) misinterpret the literature and fight a non-existing theoretical position.

A 1- $\omega$  upper bound on advance planning would anyhow be most unlikely, given the abundance of content words encompassing two or more phonological words, such as “landowner”. If the speaker would first prepare  $\omega_1$  and utter it, then prepare  $\omega_2$  and utter it, both would receive primary stress: “lánd-ówner”. But compound words in English and Dutch carry just one primary stress. In Dutch, adjectival compounds, such as “doof-stom” (deaf-mute) are stressed on the second constituent: “doof-stóm”. If the speaker would first prepare  $\omega_1$  and utter it, it would get normal primary stress, but it shouldn’t. Hence, the encoding of  $\omega_1$

(whether or not it will receive primary stress) depends on the “preview” of  $\omega_2$ . For correct stress assignment phonological encoding must work here on a 2- $\omega$  frame, although each  $\omega$  forms a separate domain of syllabification.

The case is different, and less obvious, for multiple content word utterances. Levelt and Meyer (2000) reviewed the picture naming and eye-tracking evidence for accessing the phonological code of a second content word before utterance onset. At the moment of publication no strong evidence could be reported, neither for the adjective-noun case, nor for coordination constructions, such as “the baby and the dog”. It was and is, therefore, an important issue indeed to further test under what conditions such an effect can arise. Almost all of the experimental studies had used auditory priming to affect the speed of phonological encoding of a second content word in the target utterance (for instance, presenting the subject with the auditory distractor word “doll” simultaneously with the baby-and-dog picture to be described). The first positive result with that method was recently obtained by Costa and Caramazza (2002) and by Jescheniak, Schriefers and Hantsch (in press), both testing phonological noun priming in determiner–adjective–noun phrases. It is certainly an excellent idea to cross-validate these initial findings by means of another method, varying the word frequency of a non-initial content word. That is what Alario et al. (2002) set out to do.

## DESIGNING PICTURE NAMING STUDIES OF WORD FREQUENCY AND AGE OF ACQUISITION

In both their experiments Alario et al. (2002) imposed their independent word frequency variables as follows. To control adjective word frequency they used colours that have high- versus low-frequency names. In Experiment 1 this involved two sets of four colours, in Experiment 2 two sets of two colours. It is inevitable that, for such small sets, there is some confounding with other variables, such as colour word length. As the authors argue, it is unlikely that word length has played a noticeable role. Not discussed is the issue of colour perception; recognition latencies may vary among colours and this is not controlled for in the experiment. I will, however, concentrate on the manipulation of noun frequency. That is the crucial variable to argue for a 1- $\omega$  or 2- $\omega$  planning window.

In both experiments the authors contrasted sets of depicted objects with high- versus low-frequency names, two sets of 16 pictures in Experiment 1 and two sets of 25 pictures in Experiment 2. In both cases the frequency contrast was substantial. In addition, the authors had their subjects rate the target words in terms of age-of-acquisition, on a seven-point scale. It is a

continuing matter of discussion whether the word frequency effect is in part or even entirely an age of acquisition effect (Barry, Hirsh, Johnston, & Williams, 2001; Bonin, Fayol, & Chalard, 2001). Also on this scaling measure (whether or not it reflects real age-of-acquisition—which should be a matter of debate) the two sets contrasted sharply. Excellent so far. However, the authors did not control for the speed of object recognition and this is a serious omission. Assume that, on average, the pictures with low-frequency names happen to be less recognisable than those with high-frequency names, then any frequency effect obtained in the experiment may just signal a visual process instead of a lexical one. The same data pattern may arise without there being any effect of noun frequency. This control is absolutely essential, because it may well be the case that objects with infrequent names are also less often encountered and consequently less recognisable than objects with high-frequency names.

There are various methods to impose the control, all with their advantages and disadvantages. One is to use Kroll & Potter's (1984) object/non-object decision paradigm (Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Meyer, Sleiderink, & Levelt, 1998). The target pictures are mixed with pictures of non-objects (carefully matched in other respects) and the subjects judge, by means of a push-button response, whether what they see is a really existing object or not. This decision most probably involves object recognition (just as lexical decision involves word recognition). The word frequency variable is then realised in such a way that object decision times are, on average, equal for the high- and low-frequency picture sets. Another method is verification. The subject is auditorily (Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991) or visually (Jescheniak & Levelt, 1994) presented with an object name after which a picture is presented for verification. Push button latencies are the dependent measure. Notice that Wingfield (1968) already basically used this method to validate Oldfield & Wingfield (1965) very first picture naming study of word frequency. Subjects were asked to push a button every time they saw a particular target object, say a basket, and to withhold responding otherwise. This was done for all the target words of the picture naming experiment. One drawback of this method in all its variants is that in case of "yes" responses the stimulus or target word may have facilitated recognition of the picture. The better approach is therefore to design the recognition experiment in such a way that the critical test pictures are always verified as "no", i.e., they get a mismatching word stimulus (Jescheniak & Levelt, 1994). Still another method is object recognition (first used as a control in Levelt et al., 1991, Expt. 4). Here, in a first run, a set of non-target pictures are shown to the subjects. The instruction is to remember them. Then, in a second run, the same pictures are presented, mixed with the real test pictures. For each picture the

subject pushes an “old” or a “new” button. All test pictures are new; “new” push button latencies reflect recognition speed.

The omission of recognition tests in picture naming studies of word frequency is perennial. I mention two recent examples. In their study of age-of-acquisition (AoA) versus word frequency (WF) as determiners of naming speed, Bonin et al. (2001) used picture naming to test the effect of WF if AoA ratings are held constant and inversely. They found that naming latencies are exclusively related to the AoA measure; WF plays no role if AoA is controlled for. However, they did not test their pictures for recognition speed. There was control of several other variables, such as name agreement, image agreement, visual complexity, familiarity and image variability (all reported for these pictures in Alario & Ferrand, 1999), but these were all scaling data. No tests of recognition speed were performed for the target pictures. It is not far-fetched to consider the possibility that a subject’s judgement of AoA is correlated with the speed of recognising the corresponding object. But if recognition speed did do the work in these experiments, then they do not tell us anything about lexical access. The same critique applies to a study by Barry et al. (2001). Their purpose was the same, testing the relative effects of AoA and WF in a picture naming task. In addition they studied the effect on naming latencies of having seen either the picture or the picture’s name before. They found a substantial effect of AoA on picture naming latencies if WF was controlled for. Also, they found substantial repetition priming, in particular from first to second picture naming; this priming was strongest for low AoA items. On the other hand, there was no WF effect when AoA was controlled for, completely in line with the Bonin et al. findings. However, there was no control of recognition speed in these experiments. If AoA measures are correlated with the speed of recognising the corresponding objects, which is certainly possible, if not likely, then these findings could be wholly or partially due to object recognition instead of lexical access. Luckily, Barry et al. (2001) also found an AoA effect in a plain word naming task. Here picture recognition could not have played a role. However, the effect was substantially smaller here than in the picture naming task, which increases the likelihood that recognition speed did play a substantial role in their experiments. The sad fact is that both studies are inconclusive, just as inconclusive as the Alario et al. (2002) study.

## CONCLUSION

This commentary on Alario et al. (2002) addressed two issues. The first one is that the authors critically address a theoretical position that is not held by the authors referred to. The second and more important issue concerns



the absence of a critical experimental control. Standards of experimental control develop gradually with the general progress in a discipline. It is discomfoting to see a standard set by Wingfield (1968) for the study of frequency effects in picture naming still not being respected one third of a century later.

Manuscript received November 2001  
Revised manuscript received June 2002

## REFERENCES

- Alario, F.-X., Costa, A., & Caramazza (2002). Frequency effects in Noun Phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299–319.
- Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavioral Research Methods, Instruments, & Computers*, *31*, 531–552.
- Barry, C., Hirsh, K.W., Johnston, R.A., & Williams, C.L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language*, *44*, 350–375.
- Bonin, P., Fayol, M., & Chalard, M. (2001). Age of acquisition and word frequency in written picture naming. *Quarterly Journal of Experimental Psychology*, *54A*, 469–489.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Oxford University Press.
- Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding during speech production. *Journal of Memory and Language*, *46*, 178–198.
- Jescheniak, J.D. & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.
- Jescheniak, J.D., Schriefers, H., & Hantsch, A. (in press). Utterance format affects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*.
- Kroll, J.F., & Potter, M.C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, *23*, 30–66.
- Levelt, W.J.M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M., & Meyer, A.S. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, *12*, 433–445.
- Levelt, W.J.M., Praamstra, P., Meyer, A.S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*, 553–567.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.
- Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, Th., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122–142.
- Meyer, A.S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *25*, 477–496.
- Meyer, A.S., Sleiderink, A.M., & Levelt, W.J.M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25–B33.

- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris Publications.
- Oldfield, R.C. & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.
- Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, *11*, 17–50.
- Sternberg, S., Knoll, R.L., Monsell, S., & Wright, C.E. (1988). Motor programs and the hierarchical organization in the control of rapid speech. *Phonetica*, *45*, 175–197.
- Sternberg, S., Monsell, S., Knoll, R.L., & Wright, C.E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In G.E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.
- Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381.
- Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, *81*, 226–234.