Published in final edited form as:

Cognition. 2001 November ; 82(1): B1–14.

Gaze durations during speech reflect word selection and phonological encoding

Zenzi M. Griffin

Stanford University

Abstract

Speakers produced the sentence frame *The A and the B are above the C* to describe three pictured objects while their eye movements were monitored. Object B or C varied in codability (the number of alternative names for it) and in the frequency of its dominant name. Codability is known to affect speed of word selection, and word frequency, speed to retrieve a word's pronunciation (phonological encoding). Speakers gazed longer at lower codability and lower frequency objects before naming them. However, the codability and frequency of B and C did not affect when speakers began naming A, even when utterances were perfectly fluent. The results indicate that speakers began "The A..." once they had a name prepared for A, before selecting names for B and C. Similar gaze patterns during less constrained scene description tasks in other studies suggest that speakers often incrementally select and phonologically encode nouns in fluent utterances.

Keywords

language production; word choice; time course; eye movements; incremental speech; disfluency

People often have the intuition that they speak without much preparation. But according to several models of sentence production (Garrett, 1975, 1982; Stemmerger, 1985), they decide on the message, build the syntactic structure, and select words all before they begin to speak. Only later do they prepare the pronunciations of words—what is called *phonological encoding*. According to other models (e.g., Bock, 1982; De Smedt & Kempen, 1987; Ferreira, 2000; Ward, 1992; Yngve, 1960), speakers also leave other processes to later. For instance, speakers select words incrementally (Kempen & Hoenkamp, 1987; Levelt, 1989). That is, they don't specify words at all until just before they are produced. Combining aspects of both models, Goldman-Eisler (1968) proposed that speakers are fluent when they decide on a message, structure, and particular words in advance, retrieving only the sounds on the fly. They speak hesitantly when they fail to prepare higher levels in advance.

When people describe pictured scenes such as a mailman being chased by a dog, they gaze at the mailman and dog during the second preceding the words (Griffin & Bock, 2000). This holds for both fluent and hesitant descriptions. Their gaze durations (700-900 ms) are similar to the time it takes people to name objects in isolation. When speakers describe more

complex scenes or answer questions about them, their gazes also anticipate their words by about a second (Griffin, 2000). However, people only gaze at objects for this long when performing tasks that require retrieving names (e.g., Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998; Zelinsky & Murphy, 2000). So, according to Griffin and Bock, the timing of gazes during fluent and hesitant speech reflects both incremental word selection and phonological encoding, counter to the claims of several models.

It is already known that speakers gaze longer at isolated objects before they name them when the phonological forms of the names are less accessible (Meyer et al., 1998; Meyer & van der Meulen, 2000). What remains to be shown is whether speakers in the *middle* of a description gaze longer at objects whose names are more difficult to select or take longer to encode (From here on, by “encoding a word” I will mean “encoding a word phonologically.”) The purpose of the current study was to test whether speakers’ gaze durations during speech vary with word selection difficulty. To the extent that gaze durations do, they will permit detailed inferences about the time course of word selection during sentence production.

Word selection was manipulated by varying name agreement, or *codability*. Speakers almost always call an apple *apple*, but they might call a television set a *television* or *TV*, and a schooner a *ship*, *boat*, or *sailboat*. The apple is the most codable, and the schooner the least. Speakers tend to name objects faster the more codable they are (Lachman, 1973; Lachman & Lachman, 1980). It isn't that people take more time to recognize an object of lower codability (Johnson, 1992), but that they take longer to select a name. Naming an object that has multiple names is much like naming an object in the presence of a related word. People take longer to name an object, such as a lion, when they are simultaneously presented with a semantically related word, such as *tiger* (e.g., Lupker, 1979; Schriefers, Meyer, & Levelt, 1990). Yet it doesn't take them longer to recognize the picture as an animal or as an image of a real object in the presence of the word *tiger* (Lupker, 1988; Schriefers et al, 1990). According to most theories of word production (see Levelt, 1999), words from the same category compete with each other for selection even though they are not close enough in meaning to label the same objects. Two words that can label the same object should compete with each other even more.

The difficulty of phonological encoding was manipulated by varying word frequency. People produce high frequency words faster and with fewer slips of the tongue than they do low frequency words (Oldfield & Wingfield, 1968; Stemberger & MacWhinney, 1986). It is the frequency of the word form that counts (Dell, 1990). *Ball*, for example, has two main senses—globe and dance. Speakers produce *ball* to describe dance as quickly as a word as common as *ball* in both its senses (Jescheniak & Levelt, 1994).

In the experiment reported here, speakers described displays of three objects with the sentence, *The A and the B are above the C*. One of the objects (in position B or C) varied in codability and the frequency of its dominant name. When people name these *critical objects*, they should be faster to select names for highly codable objects than for medium codable ones, regardless of the frequency of their names. Also, people should be faster to encode

high frequency names than low frequency ones, regardless of the codability of their objects¹. An example display is shown in Figure 1.

Speakers in these circumstances could prepare their descriptions in three ways. First, they could select all the names in a description, A, B, and C, before speaking fluently (Garrett, 1975, 1982; Goldman-Eisler, 1968). If they do, they should take longer to begin the description when either B or C is lower in codability. Second, they could select both of the names that precede the verb, A and B, before speaking, but save C until later (cf. Ferreira, 2000; Huitema, 1993; Smith & Wheeldon, 1999). If so, they should take longer to begin the description when B is lower in codability, but the codability of C should matter only after they begin speaking. Or third, they could select the names as needed (De Smedt & Kempen, 1987; Ward, 1992). If they do, they should take no longer to begin speaking when either B or C is lower in codability. By all three models, the frequency of B and C shouldn't affect when speech begins, but the frequency of A always should. Speakers' gazes on objects before and during descriptions, however, should reveal when they select names and phonologically encode them.

Method

Participants

Seventeen Stanford University students completed the experiment and received payment or credit in an introductory psychology course. The data of 16 are reported here². All were native speakers of American English, 18-30 years old, with corrected-to-normal or normal vision.

Materials and design

Displays were created with three line drawings of objects (see Figure 1). Each display contained one of 48 *first* objects in position A, one of 48 *critical* objects in position B or C, and one of 3 *repeated* objects in position B or C. Pictured objects were line drawings from Snodgrass and Vanderwart (1980), the Philadelphia naming test (Roach et al, 1996), and Huitema (1996).

The 48 first objects varied in name frequency to ensure that analyses of speech onset could detect lexical effects. Half of the objects had high frequency names; their noun lemmas had a mean spoken frequency of 70.6 (SE = 21) per million words in the Celex database (Baayen, Piepenbrock, & Gulikers, 1995). The other half had low frequency names, with a mean Celex frequency of 5.5 (1.2). All of the first objects were highly codable in spoken picture naming norms (Griffin & Huitema, 1999), 87.7% for high frequency names and 86.9% for low. Another three objects (ruler, ladder, needle) were repeated across trials to simplify the counterbalancing of picture combinations. They were chosen to be visually similar so that participants would fixate them on every presentation to identify them despite their repetition. The repeated objects were all highly codable, with 86.6% agreement, but

¹But see Griffin and Bock (1998) about the speed of selection affecting speed of phonological encoding.

²Eight participants were needed for a full counterbalance. Due to experimenter error, the replacement participant was run on a stimulus list that did not complete the counterbalancing of first object frequency across items.

had low frequency names, 7.7/million. The 48 critical objects varied in mean codability and frequency, as shown in Table 1. Naming latencies from norming studies (Griffin & Huitema, 1999) or isolated picture naming (Griffin & Bock, 1998, Experiment 1) were available for all but one of the objects. Analysis of item means for naming latencies showed significant, additive effects of codability, $F(1,43) = 22.81$, $MSe = 38157$, and frequency, $F(1,43) = 10.70$, $MSe = 38157$.

Data from another norming study further suggested that differences in naming latencies across conditions were not due to differences in the time needed to identify objects. Twenty Stanford students were asked to judge quickly and accurately whether each picture represented an object that was typically larger than a 8 by 11 inch sheet of paper. As expected, mean item latencies to make size judgements correlated highly with mean naming latencies, $r(n = 48) = .42$, $p < .005$. However, there were no main effects or interactions of codability and frequency, $minFs < 1.0$.

Each participant saw 48 displays containing three objects each. Critical objects formed 12 quadruples containing one object from each cell of the codability by frequency crossing. Eight stimulus lists were created to counterbalance the frequency and identity of the first object for each member in a quadruple across lists, and the position of the critical object (B or C). Across lists, two high and two low frequency first objects rotated in a Latin square through displays for a quadruple. Members of a quadruple always appeared with the same repeated object. Each list contained an equal number of items in each condition. A list of critical objects appears in the appendix.

Within-subjects factors were the position, codability, and name frequency of the critical objects and the name frequency of the first objects. The codability and frequency of the critical objects were between-items factors. The primary dependent measures were latency to the onset of the first noun and time spent gazing at objects. Gazes extend from the onset of the first fixation in the region of an object to the first saccade leaving the region. Saccades were defined as sequences of sampled points-of-regard showing a minimum distance of 0.8 degrees of visual angle and velocity of 40 degrees per second. The onset of first object names was used to mark the beginning of speech instead of the onset of the first word *the*. Factors that delay the start of speech also delay speakers in producing the name of the first object.

Apparatus

Eye movements were monitored with a remote video-based pupil/corneal reflection system, an ISCAN ETL-400 with a high-speed upgrade sampling at 120 Hz. Eye position was calibrated with a variant of McConkie's (1981) method. A ViewSonic P815 21-inch monitor displayed stimuli. A Dell computer processed eye image data, sending uncalibrated data to a Gateway computer, which was responsible for timing, presenting stimuli, digitally recording speech, calculating and recording calibrated eye position. Speech was recorded at 12 kHz via a SoundBlaster card, using a LabTec LVA 7330 headset microphone. Participants placed their foreheads against a rest to prevent movements in depth and to keep their eyes about 81 cm from the surface of the monitor. Displayed objects subtended a maximum of 5.5° of visual angle horizontally.

Procedure

Participants were tested one at a time in a large room. They were asked to describe displays with the frame, *The A and the B are above the C*, naming the objects from left to right, top to bottom. They were led through a 9-point calibration and validation routine, 3 practice trials, 1 warm-up trial, and then the 48 experimental trials. Each trial began with a fixation point in the top center of the monitor, about 4° horizontally from positions A and B. As soon as the participants fixated within about 2.3° of this point for 800 ms, the computer would display a stimulus and begin recording data. When a participant completed a sentence, the experimenter ended the trial. There was no explicit time pressure nor was fluency mentioned.

Results and Discussion

Responses and onset times for the three nouns were transcribed by a single coder, blind to item properties. All responses were included in the first analyses, regardless of the names produced. This did not change the factors of codability and frequency, because the medium codable objects had dominant and non-dominant names of similar frequencies. Analyses of variance were conducted on untrimmed means for subject and items, then combined to produce a *minF* statistic as in Clark (1973). All reported *F*'s are significant at .05, unless otherwise indicated.

Word processing before the start of speech

Speakers prepared to name only the first object, A, before they began speaking. This was shown in the time to begin naming A. It took them 152 ms less time to name A when it had a high frequency name, 1206 ms (SE = 21) than when it had a low frequency name, 1358 ms (30), *minF*(1,51) = 6.01. In contrast, the position, codability, and frequency of the critical object did not affect the time it took speakers to begin naming A at all.

Analysis of speakers' latencies to begin speaking was supported by analysis of their eye movements. Speakers spent more time gazing at A when it had a low rather than a high frequency name, 919 (32) and 735 (23) ms, respectively, *minF*(1,50) = 8.26. Nothing else affected gazes to A. Speakers briefly gazed at B before initiating A's name, but no properties of B affected the duration of their gaze in this period. When the critical object was B, speakers gazed at it for 212 (12) ms, and when it was C, they only gazed at it for 12 (5), *minF*(1,19) = 43.37. Similarly, when the repeated object was B, speakers gazed at it for 192 (11) ms, but when it was C, only for 5 (2) ms, *minF*(1,20) = 58.82. The time that speakers spent gazing at critical and repeated objects before naming A was not affected by the codability or frequency of critical objects. These analyses suggest speakers did not select names for B and C before beginning to speak.

Word processing after speech began

The time speakers spent gazing at B and C after they began speaking reflected the difficulty of selecting and phonologically encoding their names (see Table 2). Specifically, they gazed longer at critical objects that were medium rather than highly codable, *minF*(1,56) = 10.76. They gazed longer when critical objects had low rather than high frequency names, *minF*

(1,56) = 4.90. They also gazed longer when critical objects were in position C rather than B, $\min F'(1,24) = 13.78$. There were no significant effects of the first-object's name frequency or any interactions.

Speakers' gazes reflected the difficulty of selecting and encoding a name for each object. The timing of their gazes and speech indicates that they selected and phonologically encoded names for B and C after beginning to speak rather than before. These results are incompatible with the Garrett (1975, 1982) model, in which speakers select words for an entire clause before speaking. They are also inconsistent with Smith and Wheeldon's (1999) proposal that speakers generally select words for an entire subject noun phrase (*The A and the B*) before speaking. According to all accounts, however, speakers should phonologically encode words incrementally, and they did.

No model predicted that speakers would gaze longer at C than at B. Speakers did so in part because they spent longer gazing at B than C before the onset of A's name. However, the fact that there were no more objects to name after C seemed to prolong gazes further. Gazes to B typically ended 74 (24) ms before the onset of its name. In contrast, gazes to C ended 514 (52) ms after the onset of its name.

Fluent utterances

According to models that relate fluency to degree of speech planning (Goldman-Eisler, 1968; Butterworth, 1975, 1989), many disfluencies are due to speakers selecting words incrementally. The previous analyses collapsed across fluency and therefore may have been more representative of disfluent word processing than fluent.

The sentence frames that speakers used provided them with a syntactic structure and function words in advance, but they still needed to identify the particular objects and select their names for each display. If fluent utterances require that words be selected before speech begins, gazes should show codability effects before speech. To test this prediction, sentences that contained pauses over 200 ms, elongated words, or stressed articles ("thee" for *the* before a consonant; Fox Tree & Clark, 1997) were excluded, leaving 331 trials (43.1%)³. Winer's method (Winer, Brown, & Michels, 1991) was used to compensate for ten empty cells. The new analyses on completely fluent utterances showed the same pattern as the earlier ones.

The codability of critical objects had no significant effect on when speakers began to produce a name for the first object. Mean latencies and other measures are shown in Table 3. Likewise, before naming the first object, speakers spent only 34 ms longer gazing at critical objects when they were medium rather than high in codability, and the difference was not significant. It was only after beginning to name the first object that speakers' behavior was influenced by the codability of critical objects in B and C (see Figure 2). They gazed at medium codable objects for 215 ms longer than at highly codable ones. Analysis of gazes to critical objects in B shows a robust codability effect, $\min F'(1,57) = 4.47$. According to these

³Consistent with Goldman-Eisler (1968) and other's studies of fluency, sentences describing displays with medium codability objects were significantly more disfluent than others, $\min F'(1,59) = 12.43$.

results, even in the middle of a fluent utterance, a speaker's gaze reflects the time it takes to select a name for an object. Despite selecting nouns incrementally, utterances were often perfectly fluent.

Variability in word preparation

Earlier research suggested that speakers do not need to select all their words before starting a sentence (e.g., Huitema, 1993; Kempen & Huijbers, 1983; Levelt & Maassen, 1981; Lindsley, 1975; Schriefers, Teruel, & Meinhausen, 1998; Smith & Wheeldon, 1999). Their conclusions were based on speech latencies. Eye movement monitoring during descriptions provides another measure of when the processes culminating word production take place. Previous experiments have shown that gazes over figures in scenes precede their mention by about one second (Griffin, 2000; Griffin & Bock, 2000; van der Meulen & Meyer, 2000). The present experiment simply used factors related to the difficulty of word selection and phonological encoding to test which lexical processes occur during these gazes. The results indicate that these long gazes reflect the time required for both word selection and phonological encoding.

Testing effects of codability and name frequency on gaze durations limited the complexity of stimulus displays and the sentences that could be generated about them. Speakers used a sentence frame that they were given rather than generating the frame themselves. This may have encouraged incremental strategies because speakers knew in advance the structure of every sentence and where to find the objects they would name. In fact, they moved their eyes immediately to the first object to be named on 97% of trials. Without a pre-specified sentence frame, speakers typically shift their gazes a couple times before resting on the figure they choose to mention first (Griffin & Bock, 2000; van der Meulen & Meyer, 2000). When sentence structure was unpredictable, the durations of these initial, exploratory gazes were similar to those seen in non-linguistic tasks, and the estimated time for object identification (e.g., Potter, 1975). Regardless of initial picture explorations in those experiments, speakers gazed at objects during the second before naming them. The present results suggest that these one-second gazes reflect both word selection and phonological encoding in constrained and unconstrained descriptions.

Why then the variability in conclusions about the time course of word production in sentences? Speakers may prepare words further in advance when asked to (e.g., Griffin & Bock, 2000), in response to task demands (Goldman-Eisler, 1968; Huitema, 1993), and due to individual differences (e.g., Griffin & Spieler, 2000). Furthermore, when utterance-initial words are very short, a speaker may need to select the following content word before speaking fluently because there will not be a one-second interval during speech in which to encode it (Griffin, 2000). Use of objects with short names and instructions to avoid disfluencies may well have led speakers to prepare words before speech onset in experiments such as Smith and Wheeldon's (1999), but still during the second before uttering them. In addition, the preparation effects in speech onset observed in other studies may have been due to extracting an overview of the content to be expressed for structure (as van der Meulen & Meyer's data suggest) rather than due to word selection. Also, some have suggested that semantic interference effects in latency experiments (Dell & O'Seaghdha,

1992; Meyer, 1996, but see Meyer, 1997) and the occurrence of word exchanges (e.g., Garrett, 1975) may be attributed to competition or confusion in selecting a word to express part of a message rather than selection of all words before speech (Dell, 1986; Roelofs, 1998; Smith & Wheeldon, 1999). Finally, although difficulty in word selection is related to disfluency (e.g., Goldman-Eisler, 1968), incrementally selecting words does not cause speakers to hesitate. Rather, speakers are most likely to run out of preparation time when word selection is particularly difficult.

Conclusions

As they spoke, speakers' gazes reflected the difficulty of word selection and encoding for upcoming nouns. The present results go beyond the previous literature in suggesting that speakers do not even need to select every noun within an initial conjoined noun phrase before beginning to utter a sentence fluently. Together with gaze patterns observed in less constrained speech (Griffin, 2000; Griffin & Bock, 2000) and contrary to several models of production (Butterworth, 1975, 1989; Dell & O'Seaghdha, 1992; Garrett, 1982; Goldman-Eisler, 1968), the results indicate that fluent speech may result from speakers incrementally selecting their words.

Acknowledgements

Thanks to Edna Chavira, Ellen Hamilton, Amanda Silverio, and Justin Storbeck for their help in collecting and coding the data; to Kay Bock, Vic Ferreira, and the Stanford Language Users' Group for their feedback on this project; and to Herb Clark for too much to list. Funding was provided by a grant from the National Institutes of Health, R03 MH61318-01.

Appendix

Critical objects of high or medium codability and names of high or low frequency. Alternative names appear after a slash and optional elements in parentheses.

Quadruple	High-High	High-Low	Medium-High	Medium-Low
1	apple	axe	hat/cap	oven/stove
2	baby	button	board/wood	(rasp)berries
3	cake	cane	coat/jacket	crane/bulldozer
4	moon	maze	matches/matchbook	mop/broom
5	star	screw	stairs/staircase	strainer/sieve
6	window	wreath	(wine)glass	waffle iron/maker
7	table	plug	(frying)pan	palette/paints
8	bowl	butterfly	tray/platter	donkey/mule
9	bomb	well	weights/barbells	(killer)whale
10	bottle	vase	chest/trunk	vise/clamp
11	shoe	slide	(swimming) pool	sled/toboggan
12	tooth	doorknob	TV/television	limousine/limo

References

- Baayen, RH.; Piepenbrock, R.; Gulikers, L. The CELEX Lexical Database (Version 2) [CDROM]. Linguistic Data Consortium, University of Pennsylvania; Philadelphia, PA: 1995.
- Bock JK. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*. 1982; 89:1–47.
- Butterworth B. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*. 1975; 4:75–87.
- Butterworth, B. Lexical access in speech production.. In: Marslen-Wilson, WD., editor. *Lexical representation and process*. MIT Press; Cambridge, MA: 1989. p. 108-135.
- Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:335–359.
- De Smedt, K.; Kempen, G. Incremental sentence production, self-correction, and coordination.. In: Kempen, G., editor. *Natural Language Generation: New results in artificial intelligence, psychology, and linguistics*. Martinus Nijhoff; Dordrecht: 1987. p. 365-376.
- Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Dell GS. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*. 1990; 5:313–349.
- Dell GS, O'Seaghdha PG. Stages of lexical access in language production. *Cognition*. 1992; 42:287–314. [PubMed: 1582160]
- Ferreira, F. Syntax in language production: An approach using tree-adjoining grammars.. In: Wheeldon, L., editor. *Aspects of Language Production*. Psychology Press; London: 2000. p. 291-330.
- Fox Tree JE, Clark HH. Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition*. 1997; 62:151–167. [PubMed: 9141905]
- Garrett, MF. The analysis of sentence production.. In: Bower, GH., editor. *The psychology of learning and motivation*. Academic Press; New York: 1975. p. 133-177.
- Garrett, MF. Production of speech: Observations from normal and pathological language use.. In: Ellis, A., editor. *Normality and pathology in cognitive functions*. Academic Press; London: 1982. p. 19-76.
- Goldman-Eisler, F. *Psycholinguistics: Experiments in spontaneous speech*. Academic Press; London: 1968.
- Griffin, ZM. How early are verbs selected in extemporaneous English sentence production?. Paper presented at the Architectures and Mechanisms of Language Processes conference; Leiden, The Netherlands. Sep. 2000
- Griffin ZM, Bock K. Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language*. 1998; 38:313–338.
- Griffin ZM, Bock K. What the eyes say about speaking. *Psychological Science*. 2000; 11:274–279. [PubMed: 11273384]
- Griffin, ZM.; Huitema, JS. Beckman Spoken Picture Naming Norms. 1999. [On-line]. Available: <http://langprod.cogsci.uiuc.edu/~norms/>
- Griffin, ZM.; Spieler, DH. Speech planning in younger and older adults.. Poster presented at the 41st Annual Meeting of the Psychonomic Society; New Orleans, LA.. Nov. 2000
- Huitema JS. Planning referential expressions in speech production. 1993 Unpublished Dissertation, University of Massachusetts.
- Huitema JS. [The Huitema picture collection]. 1996 Unpublished line drawings.
- Jescheniak J-D, Levelt WJM. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1994; 20:824–843.
- Johnson CJ. Cognitive components of naming in children: Effects of referential uncertainty and stimulus realism. *Journal of Experimental Child Psychology*. 1992; 53:24–44. [PubMed: 1545189]

- Kempen G, Hoenkamp E. An incremental procedural grammar for sentence formulation. *Cognitive Science*. 1987; 11:201–258.
- Kempen G, Huijbers P. The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*. 1983; 14:185–209.
- Lachman R. Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*. 1973; 99:199–208.
- Lachman, R.; Lachman, JL. Picture naming: Retrieval and activation of long-term memory.. In: Poon, JLFLW.; Cermak, LS.; Grenberg, D.; Thompson, LW., editors. *New Directions in Memory and Aging*. Erlbaum; Hillsdale, NJ: 1980. p. 313-343.
- Levelt, WJM. *Speaking: From intention to articulation*. MIT Press; Cambridge, MA: 1989.
- Levelt WJM. Models of word production. *Trends in Cognitive Sciences*. 1999; 3:223–232. [PubMed: 10354575]
- Levelt, W.; Maassen, B. Lexical search and order of mention in sentence production.. In: Klein, W.; Levelt, W., editors. *Crossing the boundaries in linguistics*. Reidel; Dordrecht: 1981. p. 221-252.
- Lindsay JR. Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*. 1975; 7:1–19.
- Lupker SJ. The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*. 1979; 7:485–495.
- Lupker SJ. Picture naming: An investigation of the nature of categorical priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1988; 14:444–455.
- McConkie GW. Evaluating and reporting data quality in eye-movement research. *Behavior Research Methods & Instrumentation*. 1981; 13:97–106.
- Meyer A. Conceptual influences on grammatical planning units. *Language and Cognitive Processes*. 1997; 12:859–863.
- Meyer AS. Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*. 1996; 35:477–496.
- Meyer AS, van der Meulen FF. Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin and Review*. 2000; 7:314–319. [PubMed: 10909139]
- Meyer AS, Sleiderink A, Levelt WJM. Viewing and naming objects: Eye movements during noun phrase production. *Cognition*. 1998; 66:B25–B33. [PubMed: 9677766]
- Oldfield RC, Wingfield A. The time it takes to name an object. *Nature*. 1964; 202:1031–1032. [PubMed: 14197320]
- Potter MC. Meaning in visual search. *Science*. 1975; 187:965–966. [PubMed: 1145183]
- Roach A, Schwartz MF, Martin N, Grewal RS, Brecher A. The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*. 1996; 24:121–133.
- Roelofs A. Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1998; 24:904–921.
- Schriefers H, Meyer AS, Levelt WJM. Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*. 1990; 29:86–102.
- Schriefers H, Teruel E, Meinhausen RM. Producing simple sentence: Results from picture-word interference experiments. *Journal of Memory and Language*. 1998; 39:609–632.
- Smith M, Wheeldon L. High level processing scope in spoken sentence production. *Cognition*. 1999; 73:205–246. [PubMed: 10585515]
- Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*. 1980; 6:174–215. [PubMed: 7373248]
- Stemberger, JP. An interactive activation model of language production.. In: Ellis, A., editor. *Progress in the psychology of language*. Vol. 1. Erlbaum; London: 1985. p. 143-186.
- Stemberger JP, MacWhinney B. Frequency and the storage of regularly inflected forms. *Memory & Cognition*. 1986; 14:17–26. [PubMed: 3713503]

- van der Meulen, FF.; Meyer, AS. Coordination of eye gaze and speech in sentence production.. Poster presented at the 41st Annual Meeting of the Psychonomic Society; New Orleans, LA.. Nov. 2000
- Ward N. A parallel approach to syntax for generation. *Artificial Intelligence*. 1992; 57:183–225.
- Winer, BJ.; Brown, DR.; Michels, KM. *Statistical Principles in Experimental Design*. 3rd ed.. McGraw-Hill; New York: 1991.
- Yngve VH. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*. 1960; 104:444–466.
- Zelinsky GJ, Murphy GL. Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*. 2000; 11:125–131. [PubMed: 11273419]

Fig. 1. Example display with a typical eye movement pattern superimposed. Circles positions indicate the fixation locations and size, duration. The top center fixation is where the validation point was presented before the display. Squares around objects indicate regions used for defining gazes. The critical object in this display is the television, which is medium codable with a high frequency name. The clock is a high frequency first object.

Fig. 2. Mean time spent gazing at critical objects before and after starting to name first objects. Bars show standard errors for mean gazing time after the first name began.

Table 1

Means (and standard errors) for properties of critical objects.

Codability	Frequency	Name agreement	Spoken frequency/mln	Segments	Naming latency	Size decision latency
High	High	94 (1.7)	107 (25.2)	4.4 (0.32)	928 (36)	978 (57)
High	Low	90 (3.2)	10 (1.8)	4.8 (0.52)	1111 (50)	876 (44)
Medium	High	42 (6.1)	120 (59.9)	5.3 (0.71)	1197 (62)	935 (51)
Medium	Low	42 (5.8)	5 (1.6)	5.4 (0.60)	1387 (73)	969 (49)

Table 2

Mean (and standard error) time gazing at critical objects after starting to name A, collapsing over fluency.

Codability & Frequency	Object position	
	B	C
High High	813 (87)	1503 (134)
High Low	1045 (77)	1762 (130)
Medium High	1284 (102)	1963 (116)
Medium Low	2211 (240)	2624 (241)

Table 3

Means (and standard errors) for fluent utterances.

Object position		
Codability	B	C
Onset of A		
High	1137 (29)	1138 (26)
Medium	1154 (31)	1190 (38)
Time gazing before onset of A		
High	206 (22)	13 (13)
Medium	271 (31)	16 (7)
Time gazing after onset of A		
High	616 (55)	1426 (105)
Medium	812 (70)	1660 (106)