Looking Up "Hard Words" for a Production Test: A Comparative Study of the NOAD, MEDAL, AHD, and MW Collegiate Dictionaries

Don McCreary
University of Georgia

We test this hypothesis: The New Oxford American Dictionary *(NOAD), MW, AHD, and MEDAL equally meet the needs of American college students when they look up a* hard word. *On a production task, writing the word in an appropriate sentence, NOAD users scored much higher than the other three groups on every* hard word, *with only one exception per user. The* Macmillan English Dictionary for Advanced Learners *(MEDAL) users scored higher than the users of the* Merriam Webster's Collegiate Dictionary, *11th Edition (MW) or users of the* American Heritage Dictionary, *2nd Edition (AHD), another collegiate desk dictionary. NOAD has several advantages over the other collegiate dictionaries, including microstructure and vocabulary coverage. Unfortunately, overall coverage of* hard words *is problematic in MEDAL, since it is intended for non-natives. MW users were hampered by their tendency to choose the first sense in the entry, which is the oldest historical sense in MW. This also applies to AHD. This suggests that American college students might consider buying NOAD for its usability and its vocabulary coverage.*

1. Introduction

This paper will test three collegiate dictionaries sold to American college students, the *New Oxford American Dictionary, 2nd Edition* (NOAD hereafter) and the *Merriam Webster's Collegiate Dictionary, 11th Edition* (MW hereafter). Prior results from similar testing of the *American Heritage Dictionary, 2nd Edition* (AHD hereafter), a third American collegiate desk dictionary, will be examined. The aim is to determine how useful they may or may not be for today's college students. My hypothesis is that they are no more usable than MEDAL on a production task. The focus is on the microstructure of the entries of so-called "hard words", since they are challenging for native speakers and force them to look up the words. I examine how a traditional design for the microstructure can hinder students looking up "hard words" and how contemporary designs, such as those of the NOAD and the design in the *Macmillan English Dictionary for Advanced Learners* (MEDAL hereafter) can be helpful.

Publishing houses such as Merriam Webster and American Heritage continue to maintain time honored traditions in the presentation of microstructure for all entries, including "hard words". Hard words or *schwere Worter* have a history of articles in German lexicographic research: Mentrup (1984), Ballweg-Schramm (1983), and Wiegand (1998) are representative. In addition, I also looked at low frequency items on Lexis Nexis. All three of the "hardest words" in the writing task: *fey*, *simulacrum*, and *prolix*, are very low frequency items in Lexis Nexis, and the next three, invidious, aspersions, and variegated, are only slightly more frequent. The motivation for this research is that I have found consistent misuse of academic level "inkhorn terms" by student users over the past 25 years, even when they look up words in a collegiate dictionary. This led to an article in the IJL in 2002 about how Americans use the *American Heritage Dictionary* and a EURALEX presentation about a comprehension test of "hard words" with users aided by MEDAL or Merriam Webster Collegiate. This article has three sections: first, after the subjects and procedure are described, the productive test results are presented with their statistical significance; second, the control group results and the overall difficulty of the task are demonstrated; third, word by word results are discussed, detailing the strengths and weaknesses of the NOAD, MW, MEDAL, and AHD entries, highlighted by the students' tendencies when looking up five of the "hard words."

2. Subjects

In all, 90 subjects were tested. Three classes of American college students aided by a dictionary, totaling 78 in all, were tested. A control group (N=9) without a dictionary took the test.

Group	NOAD	MEDAL	MW	Control	NNS
N	25	28	25	9	3

Table 1. Numbers of subjects per group. N=87 plus 3 NNS

After looking at information about the subjects, it was clear that of the 90 tested, three non-native students were included in the group. Their results were discarded; only the native English speaking Americans' written results were scored and evaluated.

3. Procedure

The procedure required the 80 subjects aided by a dictionary to write one sentence for each target item, using the dictionary entries in the dictionary packet they were given. First they were to read the example sentences given at the beginning of the test in order to reduce the level of ambiguity in their sentences. Refer to the appendix for examples of acceptable and unacceptable sentences. The procedure, following that of McCreary (2002), and Fischer (1994), required the 80 subjects to follow the example sentences given at the beginning of the test in order to reduce the level of ambiguity in their sentences. The examples follow: "Clear: The used car salesman had an *unctuous* and insincere manner when he was pushing us to buy that 1985 Mercury Cougar for $5000. Unclear: My mother is sometimes *unctuous*". The researcher explained that the first example sentence was appropriate and the second example sentence was too vague and ambiguous to be considered acceptable for this exercise. Students were also asked to underline the line or phrase in the dictionary entry that was most helpful to them in the packet, similar to the procedure in Bogaards (1990).

If a student wrote a sentence that was unclear, demonstrating that he did not understand the meaning, even though the definition was at hand in the dictionary packet, this was marked unacceptable by the researcher. If the sentence was possibly acceptable in daily speech, as is "my professor is *unctuous*", this was also marked unacceptable since they had been instructed not to compose such sentences. If the student made it clear from the context he provided in his sentence that he understood the target word, the sentence was marked acceptable. The control group (N=9) also had these two examples but they did not get a dictionary packet. They had to write an appropriate sentence for each word with no help at all (also see McKeown, 1993). The researcher and his research assistant separately marked the written tasks turned in by the students and were consistent in their marking with an inter-rater reliability over 80%. When the two raters came to different conclusions, I looked up the phrase, clause, or sentence in Lexis Nexis and sometimes in Google to see if it or a similar clause with the target item might exist. If it did and was in a reputable source (journalism, literature), I accepted the student's usage of the item. The students in the NOAD aided, MW aided and MEDAL aided groups with the dictionary packets were also asked to underline the phrases in the entries that they perceived to be helpful and staple their answer sheets to the dictionary packets so that we could correlate this with their appropriate or inappropriate word usage.

4. Results

A chi square test was run on the overall results using SPSS. We used an alpha level of .01 as the threshold for the NOAD, MW, MEDAL, and control groups that determined significance. The results show that the differences in the means were not due to random chance.

Group	Ss	Total	Mean	Std. Dev.	χ2	p value
NOAD	25	176	7.04	1.038	10.375	p<.01
MEDAL	28	165	5.89	1.232	19.625	p<.01
MW	25	99	3.96	1.850	66.5	p<.01
Control	9	21	2.33	1.195	41.75	p<.01

Table 2. Chi Square Results

AHD entry	vicarious	invidious	poignant	aspersions	variegated
Unacceptable Score	49%	55%	64%	75%	86%

Table 3. AHD Results (207 Ss tested in 2001, which appeared in IJL (2002)

We tested AHD users on five of the eight words that NOAD, MW, and MEDAL users looked up and reported the percentages of unacceptable sentences for each item.

Results from students' use of the NOAD, MW and MEDAL packets. The scores indicate the number of students writing appropriate sentences for eight "hard" words. The maximum score for NOAD and MW is 25. The maximum for MEDAL is 28. Scores are given in descending order.

Target Item	NOAD Score	MW Score	MEDAL Score	Control Score
N	25	25	28	9
variegated	24 (96%)	22 (88%)	23 (82%)	7 (78%)
prolix	25 (100%)	21 (84%)	24 (86%)	0 (0%)
vicarious	23 (92%)	14 (56%)	25 (89%)	6 (67%)
aspersions	23 (92%)	12 (48%)	24 (86%)	3 (33%)
poignant	24 (96%)	11 (44%)	27 (96%)	4 (44%)
simulacrum	19 (76%)	9 (36%)	18 (64%)	0 (0%)
fey	17 (68%)	7 (28%)	6 (21%)	0 (0%)
invidious	21 (84%)	3 (12%)	18 (64%)	1 (11%)

Table 4. NOAD, Merriam Webster, MEDAL and Control Results

The overall score of the students using MW was 99, giving a mean of 3.96. This mean is notably higher than the control group's mean, 2.33. The overall score of the MEDAL students was 165, giving a mean of 5.89. The overall score of the NOAD students was 176, giving a mean of 7.04. This mean is markedly higher than the control group's mean and considerably higher than the MW group's mean. The overall score of the control group was 21, giving a mean of 2.33.

4.1. *Control group results*

A humorous attempt by a student in the control group follows: "If I can't have a cookie, can I at least have a *simulacrum*?" The control group (N=9), which had no dictionary help, tended to leave the sentences blank or wrote them in jest, as was the sentence above. Of the nine students who attempted *simulacrum*, not a single student wrote an acceptable sentence; four wrote unacceptable sentences and five students left the item blank. An example for *fey* follows: "She *feyed*

sickness in order to stay home from school," which reminds us of the so called "kid rule." For this item, *fey*, the nine students in the control group did not do well; five sentences were unacceptable and four students left the item blank. For the item *prolix*, most sentences were left blank, but one student opined: "*Prolix* sounds like a cool word." The overall score of the nine students was 21, giving a mean of 2.33.

4.2. *NOAD group results*

NOAD users did very well on this production test. The mean was 7.04, seven out of eight sentences acceptable. This was a significant difference from the mean of the MEDAL group, 5.89, nearly six acceptable, and of the MW group, 3.96, about four acceptable. NOAD has several strengths. First, the contemporary sense is given first in the entry, the most salient point for the look up. Second, the defining vocabulary tends to not be too academic; thus, no secondary lookups are needed. Third, collocations are frequently provided. Fourth, example sentences are provided that are well done, with only one exception,, the sentence for *fey*. Occasionally NOAD users are hampered by the lack of an example, as we see with *simulacrum* below. Four of the NOAD entries follow:

Aspersion *n*. (**usu. aspersions)** an attack on the reputation or integrity of someone or something: *I don't think anyone is **casting aspersions on** you.* >Late Middle English (denoting the sprinkling of water, esp. at baptism): from Latin *aspersion(n-), from aspergere (see* **ASPERSE***)*

Fêy *adj*. giving an impression of vague unworldliness: *his mother was a strange, **fey** woman*. having supernatural powers of clairvoyance. ■ *chiefly Scottish*. fated to die or at the point of death: *now he is **fey**, he sees his own death, and I see it too*. >Old English *fæge* (in the sense 'fated to die soon'). –**feyly** *adv*. – **feyness** *n*

Invidious *adj* (of an action or situation) likely to arouse or incur resentment or anger in others: *she'd put herself in an **invidious** position*. ■ (of a comparison or distinction) unfairly discriminating; unjust: *it seems **invidious** to make special mention of one aspect of his work*. >early 17th cent.: from Latin *invidiosus*, from *invidia* (see **ENVY**) –**invidiously** *adv*. **invidiousness** *n*.

Simulacrum *n*. (*pl*. –**lacra** or –**lacrums**) an image or representation of someone or something --an unsatisfactory imitation or substitute. > late 16th cent.: from Latin, from *simulare* (see **SIMULATE**).

Since *simulacrum* was often the item that caused the score to fall from a perfect eight to a seven, we can look at several examples. The first is, "The *simulacrum* of the Virgin Mary that appeared on my toast caused quite a stir." This is acceptable since it corresponds to how the' word is used in sources from Google and Lexis Nexis. Two more acceptable sentences follow: "The sculpture of my wife was definitely a simulacrum: the statue looked more like an angry bear than a beautiful woman." "George W. Bush is nothing more than a simulacrum of a true conservative." Two unacceptable attempts follow: "She acquired a negative *simulacrum* at the office after her affair with the boss was revealed." "She wouldn't drink diet sodas because she hated the horrible *simulacrums* for sugar." Occasionally, students produced a sentences with *fey* that were unacceptable: "The baby was switched by the *fey* beings." "I consider my grandmother to be very wise, a feyness about her that gives her a presence that commands respect."

4.3. *MW group results*

The students who used MW wrote sentences for *variegated* and *prolix* that tended to be appropriate. The other six items were generally taxing for the students. All of the MW users' scores were dramatically lower than the NOAD users' scores. As an example, the MW users' score for *vicarious*, 14 (56%), was encouraging, but it does not come close to the NOAD score, 23 out of 25 acceptable (92%), or the MEDAL score, 25 acceptable out of 28 attempts (89%). Comparing the MW scores with the users of the ESL dictionary, MEDAL, the scores for 5 of the 8 items of the MW users were markedly lower than the scores of the MEDAL users. The AHD users who produced sentences with *vicarious* had a score of 51% acceptable, little different from the MW score (56%).

The unacceptable sentences were usually written as NP +BE +ADJ, as in "*My professor is invidious*." This is a fine sentence in English, but the students were specifically instructed not to write this type of sentence. They were instructed to add enough details so that they would make it clear that they understood the word. The MW users often chose senses that were archaic, not realizing the senses were historically ordered. For example, when a student wrote, "I felt *poignant* when my brother punched me", this would bring a smile to well educated readers since it sounded silly. Many of the MW users' mistakes were due to this naïve use of archaic senses. The MW group had difficulty with *invidious*, *simulacrum*, *fey*, *poignant*, and *aspersions*. Four of the MW entries follow:

Aspersion *n* (ca.1587) **1:** a sprinkling with water esp. in religious ceremonies **2 a:** a false or misleading charge meant to harm someone's reputation <cast ~s on her integrity> **b:** the act of making such a charge **:** DEFAMATION

Fey *adj* **1a.** chiefly Scot: fated to die: DOOMED **b :** marked by a foreboding of death or calamity **2a:** able to see into the future: visionary **b**: marked by an otherworldly air or attitude **c:** crazy, touched **3a:** excessively refined: precious **b:** quaintly unconventional: Campy **feyly** *adv*—**feyness** *n*

Invidious *adj* [L *invidiosus* envious, invidious, fr. *invidia* envy – more at ENVY] (1606) **1:** tending to cause discontent, animosity, or envy <the ~ task of arbitration> **2:** ENVIOUS **3 a:** of

an unpleasant or objectionable nature **:** OBNOXIOUS <~ remarks> **b:** of a kind to cause harm or resentment <an ~ comparison> **- invidiously** *adv* **– invidiousness** *n*

Simulacrum *n pl,* **-cra** also **–crums** [ME, fr. L, fr. *simulare*] (15 c) **1:** IMAGE, REPRESENTATION <a reasonable ~ of reality – Martin Mayer> **2:** an insubstantial form or semblance of something: TRACE

When we examine the underlining behavior of the students, we see a tendency to underline the first sense, the oldest sense in the MW. The students' found it challenging to use this occasionally archaic sense in a sentence of their own. Students who underlined *ENVY* in the entry for *invidious* wrote sentences such as, "Sarah's new car was very *invidious* within our cluster of friends", which is unacceptable. A student who underlined "an insubstantial form" for *simulacrum* wrote, "The hallucinating girl saw a *simulacrum*". These four words above (and *prolix*) were unknown to such an extent that the students did not realize they would be better off choosing a sense other than the first sense in the entry.

4.4. *AHD group results*

If we compare the American Heritage Dictionary results from the 207 college students tested in 2001, we see that on the five words retained in the 2006-2007 testing, the sentences written by the AHD users at best had half unacceptable and at worst had 86% unacceptable. Since these results are not directly comparable, AHD cannot be ranked with the others, but we can examine each "hard word" individually. The same four entries from AHD follow:

aspersion *n.* 1. A calumnious report or remark; slander. 2. The act of defaming or slandering. 3. A sprinkling, esp. with holy water.

fey *adj*. 1. Scot. Fated to die soon. B. full of the sense of approaching death. 2. having visionary power; clairvoyant. 3. appearing as if under a spell; touched. [ME feie<OE *fæge*.]

invidious *adj*. 1. Tending to rouse ill will, animosity, or resentment; offensive. 2.

Containing or implying a slight; discriminatory. 3. Obs. Envious. [Lat. *invidiosus*, envious, hostile < *invidia*, envy 1 --see ENVY.] - invidiously *adv*. –invidiousness *n*.

simulacrum *n.*1. An image or representation of something. 2. An unreal or vague semblance of something. [Lat.<*simulare*, to simulate<*similis*, like.]

When we examine the AHD users' underlining, we see the same tendencies as the MW users who chose the first sense. The first sense is generally the oldest sense, since AHD uses historical ordering. AHD's defining vocabulary is also distinguished by its academic register, such as *calumnious*, above, which forces students into secondary look ups. It also uses abbreviations, such as *Obs*., *arch*., *ME*, and *OE*, which many students do not understand. This led to sentences such as, "I had *aspersions* of becoming an English teacher." (McCreary 2002: 198) In fact the term *aspersion* was scored 25% acceptable, a lower score than the 33% of the control group. The obsolete sense for *invidious*, *Envious*, although marked *Obs*., still led to sentences such as, "My girlfriends are very *invidious* of me because my boyfriend is handsome." (McCreary 2002: 197) If we average the AHD users' scores for the five words, we get a mean of 34% acceptable. If applied to eight words, this would yield a score of 2.72, less than the 3.96 mean of the MW users. This score is only slightly higher than the 2.33 of the control group. We can then consider some of these cases to be induced errors caused by the dictionary.

4.5. *MEDAL group results*

The MEDAL assisted students tended to write a higher number of acceptable sentences for this task for seven of the eight "hard words" than did the students who used MW. In fact, only *fey* was exceptionally challenging for MEDAL users. The same four entries from MEDAL follow:

Aspersion n. (**usu. aspersions)** an attack on the reputation or integrity of someone or something: *I don't think anyone is* **casting aspersions on you.** >Late Middle English (denoting the sprinkling of water, esp. at baptism): from Latin *aspersion(n-), from aspergere (see* **ASPERSE***)*

Fêy adj. giving the impression of vague unworldliness. His mother was a strange, fey woman. Having supernatural powers of clairvoyance. Chiefly Scottish. Fated to die or at the point of death.

Invidious adj likely to cause problems, for example by offending people or making them feel you have not treated them fairly: *invidious comparisons/choices/decisions.*

Simulacrum noun [C] very formal **1** an image that represents something **2** something that is slightly similar to something else

When we examine the MEDAL users underlining, we see the tendency to underline the first sense, which helped them since this was the contemporary sense. An exception to this pattern occurred with *invidious*. Ten of the twenty-eight underlined the second and third lines, "making them feel you have not treated them fairly," which led in every case to the production of an acceptable sentence. The entry for *simulacrum* presented a challenge for the MEDAL users. Among the more precise attempts were the following: "Prince's 'sign' is a *simulacrum* of the combination of the male and female." "*La Guernica* by Picasso is a *simulacrum* of the chaos during the Spanish Civil War". An unacceptable attempt from a MEDAL user follows: "That outfit is very *simulacrum* to the one you wore yesterday." The lack of an example or a collocation for *simulacrum* made the task difficult.

5. Conclusions

The productive tests on "hard words" for native speakers extend previously reported research on ESL and FL dictionary users by including target items that have so far been ignored by researchers. This type of research is still relatively unexplored since more user research is on ESL and bilingual dictionaries. It is obvious that productive tests on "hard words" reveal that college students are not served well by some of the American collegiate desk dictionaries. Many of the sentences produced by the MW and AHD aided groups are clear instances of induced errors. It is also evident that vocabulary coverage in learners' dictionaries, such as MEDAL, has made enormous strides.

6. Recommendations

One recommendation for Oxford is to continue doing what it has done very well in NOAD, constructing the microstructure so that the user's attention is focused on the contemporary sense of the "hard word." A minor quibble is that a few entries, such as *simulacrum* and *fey*, are in dire need of collocations and better examples. There are three recommendations for Merriam Webster and American Heritage. MW and AHD users were hampered by their tendency to choose the first sense in the entry, which is the oldest historical sense. In the preface of MW, we read, "The order of senses within an entry is historical; the sense known to have been first used in English is entered first." (Merriam-Webster 2004: 20a) I recommend that this policy be reversed. A second issue is the academic lexicon used in some of the entries. A third issue is the lack of collocations and examples. The publishers need to place the older senses at the end of the entry, decrease the number of challenging vocabulary inside the entries, and add collocations and examples.

References

Dictionaries

American Heritage Dictionary. Second College Edition. Boston: Houghton Mifflin, 1983.

Merriam-Webster's Collegiate Dictionary. 11th ed. Springfield: Merriam-Webster Incorporated, 2004.

Macmillan English Dictionary for Advanced Learners of American English. Oxford: Macmillan Education, 2002.

The New Oxford American Dictionary. 2nd ed. New York: Oxford University Press, 2005.

Other literature

Ballweg-Schramm, A. (1983). "Handbuch schwerer Wörter - Libero oder Außenverteidiger der Lexikographie?". In Henne, H.; Mentrup, W. (eds.). *Wortschatz und Verständigungs-probleme: Was sind 'schwere Wörter' im Deutschen? Sprache der Gegenwart* 57. Düsseldorf: Schwann. 143-159.

Bogaards, P. (1990). "Où cherche-t-on dans le dictionnaire?". *International Journal of Lexicography* 3 (2). 79-102.

Fischer, U. (1994). "Learning Words from Context and Dictionaries: An Experimental Comparison". *Applied Psycholinguistics* 15 (4). 551-574.

McCreary, D. R. (2002). "American Freshmen and English Dictionaries: I had *aspersions* of becoming an English Teacher". *International Journal of Lexicography* 15 (3). 181-205.

McKeown, M. G. (1993). "Creating effective definitions for young word learners". *Reading Research Quarterly* 28 (1). 16-31.

Mentrup, W. (1984). "Text, Wörter, und Wörterbuch im Deutschunterricht. Auch im Hinblich auf schwere Wörter". *Wortschatz und Wörterbuch*. *Der Deutschunterricht* 36 (5). 61-79.

Wiegand, H. E. (1998). *Wörterbuchforschung. Untersuchungen zur Wörterbuch-benutzung zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Vol. 1. Berlin: Walter de Gruyter.