Bryce Hedstrom

How Publishers Calculate Word Count

UNDERSTANDING THE "UNIQUE WORD COUNT" FORMULA

Excerpt from *Hi-Impact Reading Strategies*

When selecting books for your world language classroom library, it can be helpful to know about word counts in language learner literature. Publishers often list the unique word count and the total word count to help teachers decide on books for their classes. The unique word count is the number of different words in the book. The total word count is the length of the book.

Higher unique word counts can understandably make teachers leery about buying books, as they think the vocabulary level may be too high and therefore the story will not be comprehensible to their students. The formulas that publishers use is usually a secret, but by tedious word counts of many books and backward engineering an approximation can be formulated. The following formula results in the similar word counts as the prominent language learner publishers.

The unique word count in many independently published novels is often too high. Independent authors could often legitimately get the unique word count down by quite a bit to be in line with similar works by major publishers. Higher numbers tend to scare off some teachers because lower unique word count is shorthand for more comprehensible to their students. That makes sense.

Publishers do not want to misrepresent the number of unique words they want to keep the word count as low as honestly possible to be able to sell books and make them useful for the target readers. That also makes sense.

Here is a close approximation of how publishers calculate their unique word counts to align their products with the industry standard. We will use the beginner novel, *Pobre Ana Moderna,* by Blaine Ray as an example. Here are the principles:

- **Not every word counts.** Every word in the text of the novel should be in the glossary, but not every word in the glossary is counted for the purpose of the unique word count. It can be helpful to think in terms of the *lemma*, the stem, or the base form of a word. For example, *Pobre Ana Moderna* (see example below) has 571 words in the glossary. The book is advertised by the publisher as having a unique word count of 300. By this formula, it has 308 unique words.

- **Cognates do not count.** The cognates are highlighted in the example below. Cognates should be included in the regular alphabetical glossary for reference in case a student does not recognize them immediately, but cognates do not count in the unique word count.

- **Verbs count just one time.** Verbs fall into the lemma thinking again. Verbs are counted by infinitives or by one-time use. All verb forms in the text are included in the glossary, but each unique form of a verb is not included for the unique word count. Multiple forms of the same verb do not count. The infinitive of ***estar*** is used in the text and is counted,

but all of the other forms of it that are in the glossary and that are used in the text do not count. So, on page G-3 in the example below, these verbs forms do not count: ***está, estaba, estaban, estamos, están*** and ***estoy***. Only ***estar*** is counted. The verbs that count in the example below are the underlined ones. In some cases, only one form of a verb is used in the text and that counts.

• **Diminutives do not count.** Diminutives are not included in the unique word count, unless they are a one-time occurrence. So, if *pequeño* (small, little) and *pequeñito* (tiny, wee) are used, only *pequeño* gets counted in the unique word count. The word *pequeñito* would be included in the glossary and in the total word count, but not in the unique word count.

• **Short function words do not count.** The shortest, most common grammatical words, the function words, such as the Spanish words: ***a*** (to/at), ***y*** (and), ***o*** (or), ***e*** (and), ***u*** (or), ***el*** (the), ***la*** (the), ***le*** (to/for/at him/her/it), ***me*** (to/for/at me; myself), ***te*** (to/for/at you; yourself), ***yo*** (I), ***tú*** (you), ***al*** (at the/to the), ***en*** (in/on), ***lo*** (it), ***los*** (them), ***les*** (to/at/for them), ***del*** (of/from the), etc. are not included in the unique word count, the thinking apparently being that a student that can read this book (mid-level 1 / level 2) already knows these words and they do not greatly impact comprehensibility. The short word guideline does not apply to short verbs or other parts of speech. For beginners, these short grammatical words are not as important for conveying meaning like verbs, nouns and adjectives are, and many newbies just skip over them as they read and listen. These function words are acquired naturally as students read and listen for meaning. The function words should be included in the glossary.

In the example below, the words ***con*** (with)***, de*** (from, of)***, él*** (he) and ***el*** (the) on pages G-2 and G-3 do not count according to this unique word count formula in the new *Pobre Ana Moderna*. Those have penciled-in boxes around them.

• **Glossed words do not count.** Glossed words on pages in the text do not count because those are already acknowledged as being above the reading level of the rest of the text by virtue of the glossing. There is also no need to put the glossed words in the glossary. Subtracting the glossed words from the text often changes the unique word count substantially.

• **Proper nouns do not count.** The names of people and places do not count. The names of characters in the story, stores, products, commercial devices, cities, neighborhoods, states and countries are not counted in the unique word count either. Some of these might appear in the glossary (the name of a state of Mexico for clarification, for instance), but those words would not be included in the unique word count.

Using the guidelines above, *Pobre Ana Moderna* has a unique word count of **308 words**. The glossary in that book contains **571 words** (presumably the complete set of unique words in the text). The publisher advertises this novel as having **300 words**. The advertised unique word count is about half of the unique word count in the glossary. The publisher's formula for

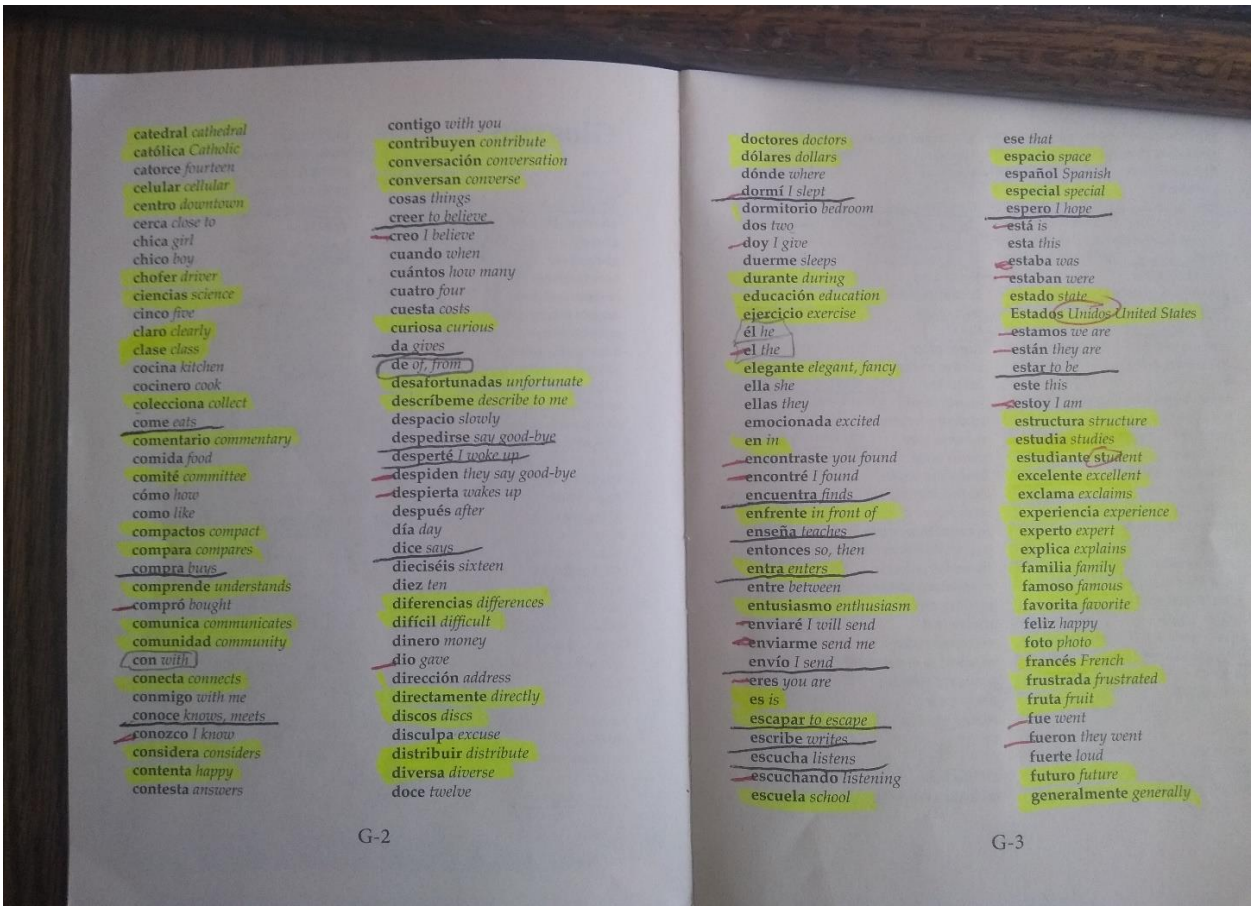
unique word counts is not public, but this backward engineering must be close to what they, and other publishers are doing.

It makes sense for a publisher to count the words low this way. Higher unique word counts have been associated with level 3 and 4 novels and they lead level 1 and 2 teachers to believe that their students will not be able to read these books without great difficulty.

Level 1 and 2 novels generally have a unique word count of 300 or fewer words.

The unique word count can reasonably be lowered using these guidelines. Publishers are not trying to trick anyone here. They just want to use the same standard that others in the industry use in order to get quality books into the hands of teachers and students. That does not tend to happen with unusually high word counts. High unique word counts do not sell because teachers think their students will not be able to understand the text.

Some pages from the glossary of Blaine Ray's best-selling novel *Pobre Ana Moderna* appear below. This is the re-write and update of the classic *Pobre Ana*, which is a good measuring stick because it was one of the first lower-level language learner novels of the modern era. The original *Pobre Ana* has sold over 1 million copies and is something of a standard against which other level 1 novels are measured for unique word count (300) and total word count (6,000), if not for story quality.



catedral *cathedral*
católica *Catholic*
catorce *fourteen*
celular *cellular*
centro *downtown*
cerca *close to*
chica *girl*
chico *boy*
chofer *driver*
ciencias *science*
cinco *five*
claro *clearly*
clase *class*
cocina *kitchen*
cocinero *cook*
colecciona *collect*
come *eats*
comentario *commentary*
comida *food*
comité *committee*
cómo *how*
como *like*
compactos *compact*
compara *compares*
compra *buys*
comprende *understands*
compró *bought*
comunica *communicates*
comunidad *community*
con *with*
conecta *connects*
conmigo *with me*
conoce *knows, meets*
conozco *I know*
considera *considers*
contenta *happy*
contesta *answers*

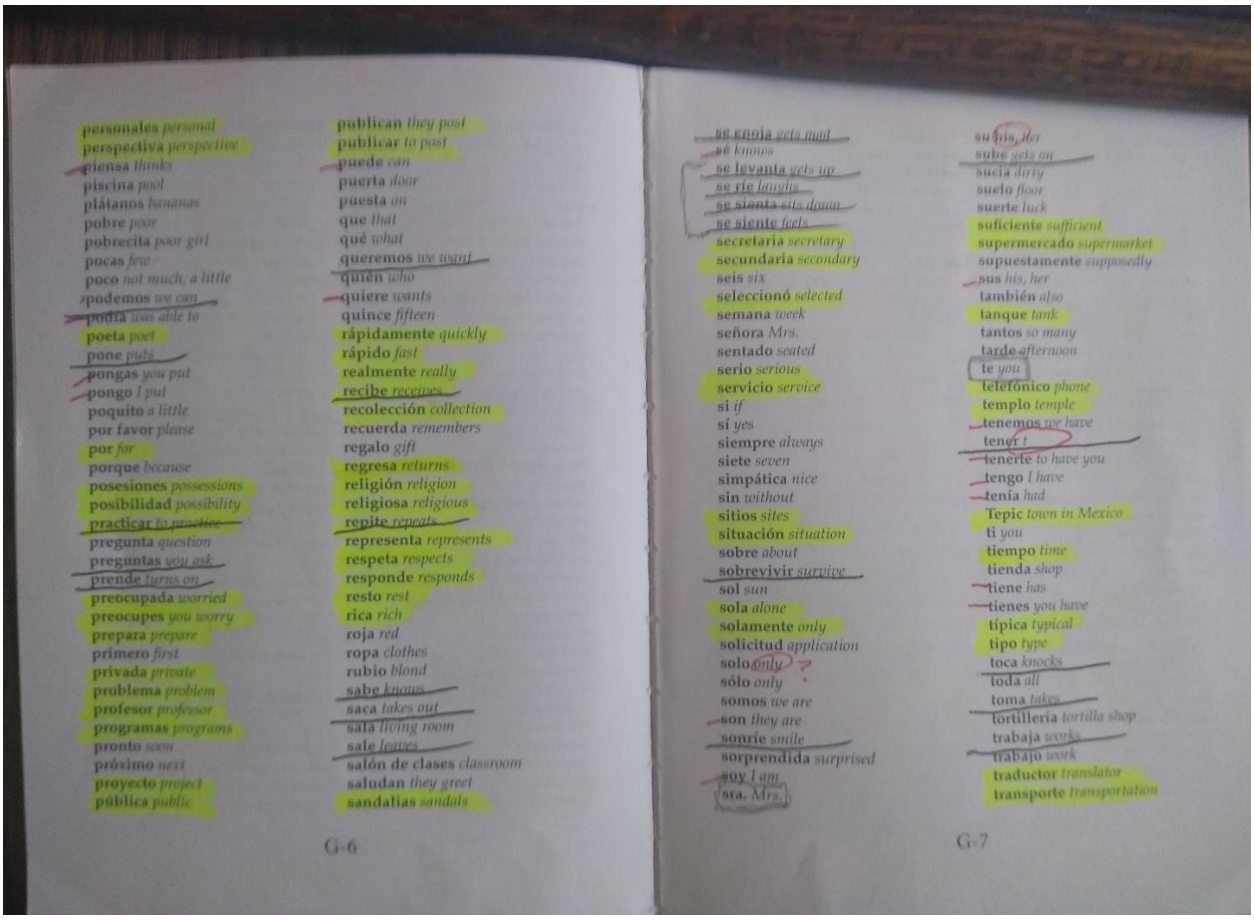
contigo *with you*
contribuyen *contribute*
conversación *conversation*
conversan *converse*
cosas *things*
creer *to believe*
creo *I believe*
cuando *when*
cuántos *how many*
cuatro *four*
cuesta *costs*
curiosa *curious*
da *gives*
de *of, from*
desafortunadas *unfortunate*
descríbeme *describe to me*
despacio *slowly*
despedirse *say good-bye*
desperté *I woke up*
despiden *they say good-bye*
despierta *wakes up*
después *after*
día *day*
dice *says*
dieciséis *sixteen*
diez *ten*
diferencias *differences*
difícil *difficult*
dinero *money*
dio *gave*
dirección *address*
directamente *directly*
discos *discs*
disculpa *excuse*
distribuir *distribute*
diversa *diverse*
doce *twelve*

G-2

doctores *doctors*
dólares *dollars*
dónde *where*
dormí *I slept*
dormitorio *bedroom*
dos *two*
doy *I give*
duerme *sleeps*
durante *during*
educación *education*
ejercicio *exercise*
él *he*
el *the*
elegante *elegant, fancy*
ella *she*
ellas *they*
emocionada *excited*
en *in*
encontraste *you found*
encontré *I found*
encuentra *finds*
enfrente *in front of*
enseña *teaches*
entonces *so, then*
entra *enters*
entre *between*
entusiasmo *enthusiasm*
enviaré *I will send*
enviarme *send me*
envío *I send*
eres *you are*
es *is*
escapar *to escape*
escribe *writes*
escucha *listens*
escuchando *listening*
escuela *school*

ese *that*
espacio *space*
español *Spanish*
especial *special*
espero *I hope*
está *is*
esta *this*
estaba *was*
estaban *were*
estado *state*
Estados Unidos *United States*
estamos *we are*
están *they are*
estar *to be*
este *this*
estoy *I am*
estructura *structure*
estudia *studies*
estudiante *student*
excelente *excellent*
exclama *exclaims*
experiencia *experience*
experto *expert*
explica *explains*
familia *family*
famoso *famous*
favorita *favorite*
feliz *happy*
foto *photo*
francés *French*
frustrada *frustrated*
fruta *fruit*
fue *went*
fueron *they went*
fuerte *loud*
futuro *future*
generalmente *generally*

G-3



personales *personal*
perspectiva *perspective*
piensa *thinks*
piscina *pool*
plátanos *bananas*
pobre *poor*
pobrecita *poor girl*
pocas *few*
poco *not much, a little*
podemos *we can*
podía *was able to*
poeta *poet*
pone *puts*
pongas *you put*
pongo *I put*
poquito *a little*
por favor *please*
por *for*
porque *because*
posesiones *possessions*
posibilidad *possibility*
practicar *to practice*
pregunta *question*
preguntas *you ask*
prende *turns on*
preocupada *worried*
preocupes *you worry*
prepara *prepare*
primero *first*
privada *private*
problema *problem*
profesor *professor*
programas *programs*
pronto *soon*
próximo *next*
proyecto *project*
pública *public*

publican *they post*
publicar *to post*
puede *can*
puerta *door*
puesta *on*
que *that*
qué *what*
queremos *we want*
quién *who*
quiere *wants*
quince *fifteen*
rápidamente *quickly*
rápido *fast*
realmente *really*
recibe *receives*
recolección *collection*
recuerda *remembers*
regalo *gift*
regresa *returns*
religión *religion*
religiosa *religious*
repite *repeats*
representa *represents*
respeta *respects*
responde *responds*
resto *rest*
rica *rich*
roja *red*
ropa *clothes*
rubio *blond*
sabe *knows*
saca *takes out*
sala *living room*
sale *leaves*
salón de clases *classroom*
saludan *they greet*
sandalias *sandals*

G-6

se enoja *gets mad*
sé *knows*
se levanta *gets up*
se ríe *laughs*
se sienta *sits down*
se siente *feels*
secretaria *secretary*
secundaria *secondary*
seis *six*
seleccionó *selected*
semana *week*
señora *Mrs.*
sentado *seated*
serio *serious*
servicio *service*
si *if*
sí *yes*
siempre *always*
siete *seven*
simpática *nice*
sin *without*
sitios *sites*
situación *situation*
sobre *about*
sobrevivir *survive*
sol *sun*
sola *alone*
solamente *only*
solicitud *application*
solo *only*
sólo *only*
somos *we are*
son *they are*
sonríe *smile*
sorprendida *surprised*
soy *I am*
sra. *Mrs.*

su *his, her*
sube *gets on*
sucia *dirty*
suelo *floor*
suerte *luck*
suficiente *sufficient*
supermercado *supermarket*
supuestamente *supposedly*
sus *his, her*
también *also*
tanque *tank*
tantos *so many*
tarde *afternoon*
te *you*
telefónico *phone*
templo *temple*
tenemos *we have*
tener *t*
tenerte *to have you*
tengo *I have*
tenía *had*
Tepic *town in Mexico*
ti *you*
tiempo *time*
tienda *shop*
tiene *has*
tienes *you have*
típica *typical*
tipo *type*
toca *knocks*
toda *all*
toma *takes*
tortillería *tortilla shop*
trabaja *works*
trabajo *work*
traductor *translator*
transporte *transportation*

G-7