

Word Maturity: Computational Modeling of Word Knowledge

Kirill Kireyev

Thomas K Landauer

Pearson Education, Knowledge Technologies
Boulder, CO

{kirill.kireyev, tom.landauer}@pearson.com

Abstract

While computational estimation of difficulty of words in the lexicon is useful in many educational and assessment applications, the concept of scalar word difficulty and current corpus-based methods for its estimation are inadequate. We propose a new paradigm called *word meaning maturity* which tracks the degree of knowledge of each word at different stages of language learning. We present a computational algorithm for estimating word maturity, based on modeling language acquisition with Latent Semantic Analysis. We demonstrate that the resulting metric not only correlates well with external indicators, but captures deeper semantic effects in language.

1 Motivation

It is no surprise that through stages of language learning, different words are learned at different times and are known to different extents. For example, a common word like “dog” is familiar to even a first-grader, whereas a more advanced word like “focal” does not usually enter learners’ vocabulary until much later. Although individual rates of learning words may vary between high- and low-performing students, it has been observed that “children [...] acquire word meanings in roughly the same sequence” (Biemiller, 2008).

The aim of this work is to model the degree of knowledge of words at different learning stages. Such a metric would have extremely useful applications in personalized educational technologies, for the purposes of accurate assessment and personalized vocabulary instruction.

2 Rethinking Word Difficulty

Previously, related work in education and psychometrics has been concerned with measuring *word difficulty* or classifying words into different difficulty categories.

Examples of such approaches include creation of word lists for targeted vocabulary instruction at various grade levels that were compiled by educational experts, such as Nation (1993) or Biemiller (2008). Such word difficulty assignments are also implicitly present in some readability formulas that estimate difficulty of texts, such as Lexiles (Stenner, 1996), which include a lexical difficulty component based on the frequency of occurrence of words in a representative corpus, on the assumption that word difficulty is inversely correlated to corpus frequency. Additionally, research in psycholinguistics has attempted to outline and measure psycholinguistic dimensions of words such as *age-of-acquisition* and *familiarity*, which aim to track when certain words become known and how familiar they appear to an average person.

Importantly, all such word *difficulty* measures can be thought of as functions that assign a single scalar value to each word w :

$$\text{difficulty} : f(w) \rightarrow \mathbb{R} \quad (1)$$

There are several important limitations to such metrics, regardless of whether they are derived from corpus frequency, expert judgments or other measures.

First, learning each word is a continual process, one that is interdependent with the rest of the vocabulary. Wolter (2001) writes:

[...] Knowing a word is quite often not an either-or situation; some words are known well, some not at all, and some are known to varying degrees. [...] How well a particular word is known may condition the connections made between that particular word and the other words in the mental lexicon.

Thus, instead of modeling *when* a particular word will become fully known, it makes more sense to model the *degree* to which a word is known at different levels of language exposure.

Second, word difficulty is inherently perspectival: the degree of word understanding depends not only on the word itself, but also on the sophistication of a given learner. Consider again the difference between “dog” and “focal”: a typical first-grader will have much more difficulty understanding the latter word compared to the former, whereas a well-educated adult will be able to use these words with equal ease. Therefore, the degree, or *maturity*, of word knowledge is inherently a function of two parameters -- word w and learner level l :

$$maturity : f(w, l) \rightarrow \mathbb{R} \quad (2)$$

As the level l increases (i.e. for more advanced learners), we would expect the degree of understanding of word w to approach its full value corresponding to perfect knowledge; this will happen at different rates for different words.

Ideally, we would obtain maturity values by testing word knowledge of learners across different levels (ages or school grades) for all the words in the lexicon. Such a procedure, however, is prohibitively expensive; so instead we would like to estimate word maturity by using computational models.

To summarize: our aim is to model the development of *meaning* of words as a function of increasing exposure to language, and ultimately - the degree to which the meaning of words at each stage of exposure resemble their “adult” meaning. We therefore define *word meaning maturity* to be the degree to which the understanding of the word (expected for the average learner of a particular level) resembles that of an ideal mature learner.

3 Modeling Word Meaning Acquisition with Latent Semantic Analysis

3.1 Latent Semantic Analysis (LSA)

An appealing choice for quantitatively modeling word meanings and their growth over time is Latent Semantic Analysis (LSA), an unsupervised method for representing word and document meaning in a multi-dimensional vector space.

The LSA vector representation is derived in an unsupervised manner, based on occurrence patterns of words in a large corpus of natural language documents. A Singular Value Decomposition on the high-dimensional matrix of word/document occurrence counts (A) in the corpus, followed by zeroing all but the largest r elements¹ of the diagonal matrix S , yields a lower-rank word vector matrix (U). The dimensionality reduction has the effect of smoothing out incidental co-occurrences and preserving significant semantic relationships between words. The resulting word vectors² in U are positioned in such a way that semantically related words vectors point in similar directions or, equivalently, have higher cosine values between them. For more details, please refer to Landauer et al. (2007) and others.

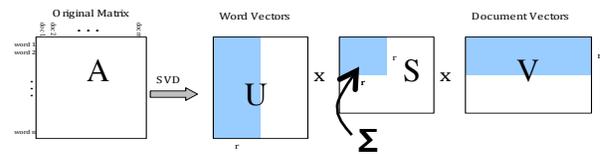


Figure 1. The SVD process in LSA illustrated. The original high-dimensional word-by-document matrix A is decomposed into word (U) and document (V) matrices of lower dimensionality.

In addition to merely measuring semantic relatedness, LSA has been shown to emulate the learning of word meanings from natural language (as can be evidenced by a broad range of applications from synonym tests to automated essay grading), at rates that resemble those of human learners (Landauer et al, 1997). Landauer and Dumais (1997) have demonstrated empirically that LSA can emulate not only the rate of human language acquisition, but also more subtle phenomena, such as the effects of learning certain words on meaning of other words. LSA can model meaning with

¹ Typically the first approx. 300 dimensions are retained

² $U\Sigma$ is used to project word vectors into V -space

high accuracy, as attested, for example, by 90% correlation with human judgments on assessing the quality of student essay content (Landauer, 2002).

3.2 Using LSA to Compute Word Maturity

In this work, the general procedure behind computationally estimating word maturity of a learner at a particular intermediate level (i.e. age or school grade level) is as follows:

1. Create an intermediate corpus for the given level. This corpus approximates the amount and sophistication of language encountered by a learner at the given level.
2. Build an LSA space on that corpus. The resulting LSA word vectors model the meaning of each word to the particular intermediate-level learner.
3. Compare the meaning representation of each word (its LSA vector) to the corresponding one in a reference model. The reference model is trained on a much larger corpus and approximates the word meanings by a mature adult learner.

We can repeat this process for each of a number of levels. These levels may directly correspond to school grades, learner ages or any other arbitrary gradations.

In summary, we estimate word maturity of a given word at a given learner level by comparing the word vector from an *intermediate* LSA model (trained on a corpus of size and sophistication comparable to that which a typical real student at the given level encounters) to the corresponding vector from a reference *adult* LSA model (trained on a larger corpus corresponding to a mature language learner). A high discrepancy between the vectors would suggest that an intermediate model's meaning of a particular word is quite different from the reference meaning, and thus the word maturity at the corresponding level is relatively low.

3.3 Procrustes Alignment (PA)

Comparing vectors across different LSA spaces is less straightforward, since the individual dimensions in LSA do not have a meaningful interpretation, and are an artifact of the content and ordering of the training corpus used. Therefore, direct com-

parisons across two different spaces, even of the same dimensionality, are meaningless, due to a mismatch in their coordinate systems.

Fortunately, we can employ a multivariate algebra technique known as Procrustes Alignment (or Procrustes Analysis) (PA) typically used to align two multivariate configurations of a corresponding set of points in two different geometric spaces. PA has been used in conjunction with LSA, for example, in cross-language information retrieval (Littman, 1998).

The basic idea behind PA is to derive a rotation matrix that allows one space to be rotated into the other. The rotation matrix is computed in such a way as to minimize the differences (namely: sum of squared distances) between corresponding points, which in the case of LSA can be common words or documents in the training set.

For more details, the reader is advised to consult chapter 5 of (Krzanowski, 2000) or similar literature on multivariate analysis. In summary, given two matrices containing coordinates of n corresponding points X and Y (and assuming mean-centering and equal number of dimensions, as is the case in this work), we would like to minimize the sum of squared distances between the points:

$$M^2 = \sum_{i=1}^n \sum_{j=1}^n (x_{ij} - y_{ij})^2$$

We try to find an orthogonal rotation matrix Q , which minimizes M^2 by rotating Y relative to X . That matrix can be obtained by solving the equation:

$$M^2 = \text{trace}(XX' + YY' - 2XQ'Y')$$

It turns out that the solution to Q is given by VU' , where $U\Sigma V'$ is the singular value decomposition of the matrix $X'Y$.

In our situation, where there are two spaces, *adult* and *intermediate*, the alignment points are the corresponding document vectors corresponding to the documents that the training corpora of the two models have in common (recall that the adult corpus is a superset of each of the intermediate corpora). The result of the Procrustes Alignment of the two spaces is effectively a joint LSA space containing two distinct word vectors for each word (e.g. “*dog1*”, “*dog2*”), corresponding to the vectors from each of the original spaces. After

merging using Procrustes Alignment, the comparison of word meanings becomes a simple problem of comparing word vectors in the joint space using the standard cosine metric.

4 Implementation Details

In our experiments we used passages from the MetaMetrics Inc. 2002 corpus³, largely consisting of educational and literary content representative of the reading material used in American schools at different grade levels. The average length of each passage is approximately 135 words.

The first-level intermediate corpus was composed of 6,000 text passages, intended for school grade 1 or below. The grade level is approximated using the Coleman-Liau readability formula (Coleman, 1975), which estimates the US grade level necessary to comprehend a given text, based on its average sentence and word length statistics:

$$CLI = 0.0588L - 0.296S - 15.8 \quad (4)$$

where L is the average number of letters per 100 words and S is the average number of sentences per 100 words.

Each subsequent intermediate corpus contains additional 6,000 new passages of the next grade level, in addition to the previous corpus. In this way, we create 14 levels. The adult corpus is twice as large, and of same grade level range (0-14) as the largest intermediate corpus.

In summary, the following describes the size and makeup of the corpora used:

Corpus	Size (passages)	Approx. Grade Level (Coleman-Liau Index)
Intermediate 1	6,000	0.0 - 1.0
Intermediate 2	12,000	0.0 - 2.0
Intermediate 3	18,000	0.0 - 3.0
Intermediate 4	24,000	0.0 - 4.0
...		
Intermediate 14	84,000	0.0 - 14.0
Adult	168,000	0.0 - 14.0

Table 1. Size and makeup of corpora. used for LSA models.

The particular choice of the Coleman-Liau readability formula (CLI) is not essential; our experiments show that other well-known readability formulas (such as Lexiles) work equally well. All that is needed is some *approximate* ordering of

³ We would like to acknowledge Jack Stenner and MetaMetrics for the use of their corpus.

passages by difficulty, in order to mimic the way typical human learners encounter progressively more difficult materials at successive school grades.

After creating the corpora, we:

1. Build LSA spaces on the adult and each of the intermediate corpora
2. Merge the intermediate space for level l with the adult space, using Procrustes Alignment. This results in a joint space with two sets of vectors: the versions from the intermediate space $\{vl_w\}$, and adult space $\{va_w\}$.
3. Compute the cosine in the joint space between the two word vectors for the given word w

$$wm(w, l) = \cos(vl_w, va_w) \quad (5)$$

In the cases where a word w has not been encountered in a given intermediate space, or in the rare cases where the cosine value falls below 0, the word maturity value is set to 0. Hence, the range for the word maturity function falls in the closed interval $[0.0, 1.0]$. A higher cosine value means greater similarity in meaning between the reference and intermediate spaces, which implies a more mature meaning of word w at the level l , i.e. higher word meaning maturity. The scores between discrete levels are interpolated, resulting in a continuous word maturity curve for each word.

Figure 1 below illustrates resulting word maturity curves for some of the words.

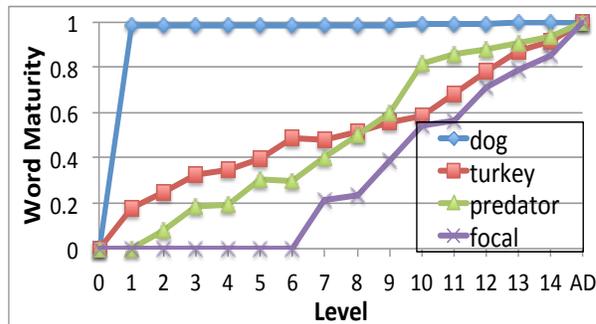


Figure 2. Word maturity curves for selected words.

Consistent with intuition, simple words like “dog” approach their adult meaning rather quickly, while “focal” takes much longer to become known to any degree.

An interesting example is “turkey”, which has a noticeable plateau in the middle. This can be explained by the fact that this word has two distinct senses. Closer analysis of the corpus and the semantic near-neighbor word vectors at each in-

intermediate space, shows that earlier meanings deal almost exclusively with the first sense (bird), while later readings with the other (country). Therefore, even though the word “turkey” is quite prevalent in earlier readings, its full meaning is not learned until later levels. This demonstrates that our method takes into account the *meaning*, and not merely the frequency of occurrence.

5 Evaluation

5.1 Time-to-maturity

Evaluation of the word maturity metric against external data is not always straightforward because, to the best of our knowledge, data that contains word knowledge statistics *at different learner levels* does not exist. Instead, we often have to evaluate against external data consisting of scalar *difficulty* values (see Section 2 for discussion) for each word, such as age-of-acquisition norms described in the following subsection.

There are two ways to make such comparisons possible. One is to compute the word maturity at a particular level, obtaining a single number for each word. Another is by computing *time-to-maturity*: the minimum level (the value on the x-axis of the word maturity graph) at which the word maturity reaches⁴ a particular threshold α :

$$ttm(w) = \min(l) \text{ s.t. } wm(w, l) > \alpha \quad (6)$$

Intuitively, this measure corresponds to the age in a learner’s development when a given word becomes sufficiently understood. The parameter α can be estimated empirically (in practice $\alpha=0.45$ gives good correlations with external measures). Since the values of word maturity are interpolated, the $ttm(w)$ can take on fractional values.

It should be emphasized that such a collapsing of word *maturity* into a scalar value inherently results in loss of information; we only perform it in order to allow evaluation against external data sources.

As a baseline for these experiments we include word frequency, namely the document frequency of words in the adult corpus.

5.2 Age-of-Acquisition Norms

Age-of-Acquisition (AoA) is a psycholinguistic property of words originally reported by Carol & White (1973). Age of Acquisition approximates the age at which a word is first learned and has been proposed as a significant contributor to language and memory processes. With some exceptions, AoA norms are collected by subjective measures, typically by asking each of a large number of participants to estimate in years the age when they have learned the word. AoA estimates have been shown to be reliable and provide a valid estimate for the objective age at which a word is acquired; see (Davis, in press) for references and discussion.

In this experiment we compute Spearman correlations between time-to-maturity and two available collections of AoA norms: Gilhooly et al., (1980) norms⁵, and Bristol norms⁶ (Stadthagen-Gonzalez et al., 2010).

Measure	Gilhooly (n=1643)	Bristol (n=1402)
(-) Frequency	0.59	0.59
Time-to-Maturity ($\alpha=0.45$)	0.72	0.64

Table 2. Correlations with Age of Acquisition norms.

5.3 Instruction Word Lists

In this experiment, we examine leveled lists of words, as created by Biemiller (2008) in the book entitled “Words Worth Teaching: Closing the Vocabulary Gap”. Based on results of multiple-choice word comprehension tests administered to students of different grades as well as expert judgments, the author derives several word difficulty lists for vocabulary instruction in schools, including:

- Words known by most children in grade 2
- Words known by 40-80% of children in grade 2
- Words known by 40-80% of children in grade 6
- Words known by fewer than 40% of children in grade 6

One would expect the words in these four groups to increase in difficulty, in the order they are presented above.

⁴ Values between discrete levels are obtained using piecewise linear interpolation

⁵ http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

⁶ http://language.psy.bris.ac.uk/bristol_norms.html

To verify how these word groups correspond to the word maturity metric, we assign each of the words in the four groups a difficulty rating 1-4 respectively, and measure the correlation with time-to-maturity.

Measure	Correlation
(-) Frequency	0.43
Time-to-maturity ($\alpha=0.45$)	0.49

Table 3. Correlations with instruction word lists (n=4176).

The word maturity metric shows higher correlation with instruction word list norms than word frequency.

5.4 Text Complexity

Another way in which our metric can be evaluated is by examining the word maturity in texts that have been leveled, i.e. have been assigned ratings of difficulty. On average, we would expect more difficult texts to contain more difficult words. Thus, the correlation between text difficulty and our word maturity metric can serve as another validation of the metric.

For this purpose, we obtained a collection of readings that are used as reading comprehension tests by different state websites in the US⁷. The collection consists of 1,220 readings, each annotated with a US school grade level (in the range between 3-12) for which the reading is intended. The average length each passage was approximately 489 words.

In this experiment we computed the correlation of the grade level with time-to-maturity, and two other measures, namely:

- *Time-to-maturity*: average time-to-maturity of unique words in text (excluding stopwords) with $\alpha=0.45$.
- *Coleman-Liau*. The Coleman-Liau readability index (Equation 4).
- *Frequency*. Average of corpus log-frequency for unique words in the text, excluding stopwords.

Measure	Correlation
Frequency (avg. of unique words)	0.60
Coleman-Liau	0.64
Time-to-maturity ($\alpha=0.45$) (avg. of unique non-stopwords)	0.70

Table 4. Correlations of grade levels with different metrics.

6 Emphasis on Meaning

In this section, we would like to highlight certain properties of the LSA-based word maturity metric, particularly aiming to illustrate the fact that the metric tracks acquisition of meaning from exposure to language and not merely more shallow effects, such as word frequency in the training corpus.

6.1 Maturity based on Frequency

For a baseline that does not take meaning into account, let us construct a set of maturity-like curves based on frequency statistics alone. More specifically, we define the *frequency-maturity* for a particular word at a given level as the ratio of the number of occurrences at the intermediate corpus for that level (l) to the number of occurrences in the reference corpus (a):

$$fm(w, l) = \frac{num_occur_l(w)}{num_occur_a(w)}$$

Similarly to the original LSA-based word maturity metric, this ratio increases from 0 to 1 for each word as the amount of cumulative language exposure increases. The corpora used at each intermediate level are identical to the original word maturity model, but instead of creating LSA spaces we simply use the corpora to compute word frequency.

The following figure shows the Spearman correlations between the external measures used for experiments in Section 5, and time-to-maturity computed based on the two maturity metrics: the new frequency-based maturity and the original LSA-based word maturity.

⁷ The collection was created as part of the “Aspects of Text Complexity” project funded by the Bill and Melinda Gates Foundation, 2010.

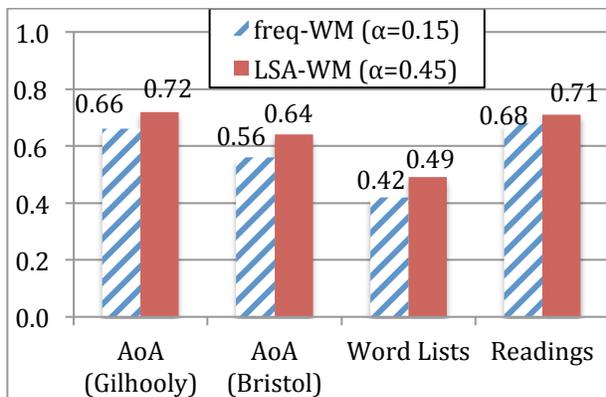


Figure 3. Correlations of word maturity computed using frequency (as well as the original) against external metrics described in Section 5.

The results indicate that the original LSA-based word maturity correlates better with real-world data than a maturity metric simply based on frequency.

6.2 Homographs

Another insight into the fact that the LSA-based word maturity metric tracks word meaning rather than mere frequency may be gained from analysis of words that are *homographs*: words that contain two or more unrelated meanings in the same written form, such as the word “turkey” illustrated in Section 4. (This is related to but distinct from the merely *polysemous* words that have several related meanings),

Because of the conflation of several unrelated meanings into the same orthographic form, homographs implicitly contain more semantic content in a single word. Therefore, one would expect the meaning of homographs to mature more slowly than would be predicted by frequency alone: all things being equal, a learner has to learn the meanings for all of the senses of a homograph word before the word can be considered fully known.

More specifically, one would expect the time-to-maturity of homographs to have greater values than words of similar frequency. To test this hypothesis, we obtained⁸ a list 174 common English homographs. For each of them, we compared their time-to-maturity to the average time-to-maturity of words that have the same (+/- 1%) corpus frequency.

⁸ http://en.wikipedia.org/wiki/List_of_English_homographs

The results of a paired t-test confirms the hypothesis that the time-to-maturity of homographs is greater than other words of the same frequency, with the p -value = $5.9e^{-6}$. This is consistent with the observation that homographs will take longer to learn and serves as evidence that LSA-based word maturity approximates effects related to meaning.

6.3 Size of the Reference Corpus

Another area of investigation is the repercussions of the choice of the corpus for the reference (adult) model. The size (and content) of the corpus used to train the reference model is potentially important, since it affects the word maturity calculations, which are comparisons of the intermediate LSA spaces to the reference LSA space built on this corpus.

It is interesting to investigate how the word maturity model would be affected if the adult corpus were made significantly more sophisticated. If the word maturity metric were simply based on word frequency (including the frequency-based maturity baseline described in Section 6.1), one would expect the word maturity of the words at each level to *decrease* significantly if the reference model is made significantly larger, since each intermediate level will have encountered fewer words by comparison. Intuition about language learning, however, tells us that with enough language exposure a learner learns virtually all there is to know about any particular word; after the word reaches its adult maturity, subsequent encounters of natural readings do little to further change the knowledge of that word. Therefore, if word maturity were tracking something similar to real word knowledge, one would expect the word maturity for most words to plateau over time, and subsequently not change significantly, no matter how sophisticated the reference model becomes.

To evaluate this inquiry we created a reference corpus that is twice as large as before (four times as large and of the same difficulty range as the corpus for the last intermediate level), containing roughly 329,000 passages. We computed the word maturity model using this larger reference corpus, while keeping all the original intermediate corpora of the same size and content.

The results show that the average word maturity of words at the last intermediate level (14) de-

creases by less than 14% as a result of doubling the adult corpus. Furthermore, this number is as low as 6%, if one only considers more common words that occur 50 times or more in the corpus. This relatively small difference, in spite of a two-fold increase of the adult corpus, is consistent with the idea that word knowledge should approach a plateau, after which further exposure to language does little to change most word meanings.

6.4 Integration into Lexicon

Another important consideration with respect to word learning mentioned in Wotler (2001), is the “connections made between [a] particular word and the other words in the mental lexicon.” One implication of that is that measuring word maturity must take into account the way words in the language are integrated with other words.

One way to test this effect is to introduce readings where a large part of the important vocabulary is not well known to learners at a given level. One would expect learning to be impeded when the learning materials are inappropriate for the learner level.

This can be simulated in the word maturity model by rearranging the order of some of the training passages, by introducing certain advanced passages at a very early level. If the results of the word maturity metric were merely based on frequency, such a reordering would have no effect on the maturity of important words (measured after all the passages containing these words have been encountered), since the total number of relevant word encounters does not change as a result of this reshuffling. If, however, the metric reflected at least some degree of semantics, we would expect word maturities for important words in these readings to be lower as a result of such rearranging, due to the fact that they are being introduced in contexts consisting of words that are not well known at the early levels.

To test this effect, we first collected all passages in the training corpus of intermediate models containing some advanced words from different topics, namely: “chromosome”, “neutron” and “filibuster” together with their plural variants. We changed the order of inclusion of these 89 passages into the intermediate models in each of the two following ways:

1. All the passages were introduced at the first level ($l=1$) intermediate corpus
2. All the passages were introduced at the last level ($l=14$) intermediate corpus.

This resulted in two new variants of word maturity models, which were computed in all the same ways as before, except that all of these 89 advanced passages were introduced either at the very first level or at the very last level. We then computed the word maturity at the levels they were introduced. The hypothesis consistent with a meaning-based maturity method would be that less learning (i.e. lower word maturity) of the relevant words will occur when passages are introduced prematurely (at level 1). Table 5 shows the word maturities measured for each of those cases, at the level (l or 14) when all of the passages have been introduced.

Word	Introduced at $l=1$ (WM at $l=1$)	Introduced at $l=14$ (WM at $l=14$)
chromosome	0.51	0.73
neutron	0.51	0.72
filibuster	0.58	0.85

Table 5. Word maturity of words resulting when all the relevant passages are introduced early vs late.

Indeed, the results show lower word maturity values when advanced passages are introduced too early, and higher ones when the passages are introduced at a later stage, when the rest of the supporting vocabulary is known.

7 Conclusion

We have introduced a new metric for estimating the degree of knowledge of words by learners at different levels. We have also proposed and evaluated an implementation of this metric using Latent Semantic Analysis.

The implementation is based on unsupervised word meaning acquisition from natural text, from corpora that resemble in volume and complexity the reading materials a typical human learner might encounter.

The metric correlates better than word frequency to a range of external measures, including vocabulary word lists, psycholinguistic norms and leveled texts. Furthermore, we have shown that the metric is based on word meaning (to the extent that it can be approximated with LSA), and not merely on shallow measures like word frequency.

Many interesting research questions still remain pertaining to the best way to select and partition the training corpora, align adult and intermediate LSA models, correlate the results with real school grade levels, as well as other free parameters in the model. Nevertheless, we have shown that LSA can be employed to usefully mimic model word knowledge. The models are currently used (at Pearson Education) to create state-of-the-art personalized vocabulary instruction and assessment tools.

References

- Andrew Biemiller (2008). *Words Worth Teaching*. Columbus, OH: SRA/McGraw-Hill.
- John B. Carroll and M. N. White (1973). Age of acquisition norms for 220 picturable nouns. *Journal of Verbal Learning & Verbal Behavior*, 12, 563-576.
- Meri Coleman and T.L. Liao (1975). *A computer readability formula designed for machine scoring*, Journal of Applied Psychology, Vol. 60, pp. 283-284.
- Ken J. Gilhooly and R. H. Logie (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods & Instrumentation*, 12, 395-427.
- Wojtek J. Krzanowski (2000) *Principles of Multivariate Analysis: A User's Perspective* (Oxford Statistical Science Series). Oxford University Press, USA.
- Thomas K Landauer and Susan Dumais (1997). A solution to Plato's problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104, pp 211-240.
- Thomas K Landauer (2002). On the Computation Basis of Learning and Cognition: Arguments from LSA. In N. Ross (Ed.), *The Psychology of Learning and Motivation*, 41, 43-84.
- Thomas K Landauer, Danielle S. McNamara, Simon Dennis, and Walter Kintsch (2007). *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum.
- Paul Nation (1993). Measuring readiness for simplified material: a test of the first 1,000 words of English. In *Simplification: Theory and Application* M. L. Tickoo (ed.), RELC Anthology Series 31: 193-203.
- Hans Stadthagen-Gonzalez and C. J. Davis (2006). The Bristol Norms for Age of Acquisition, Imageability and Familiarity. *Behavior Research Methods*, 38, 598-605.
- A. Jackson Stenner (1996). Measuring Reading Comprehension with the Lexile Framework. *Forth North American Conference on Adolescent/Adult Literacy*.
- Brent Wolter (2001). Comparing the L1 and L2 Mental Lexicon. *Studies in Second Language Acquisition*. Cambridge University Press.