

Title:	<i>WordClassifier</i> version 2.5 (March 14, 2004)	
Program Type:	Tool for lesson preparation for EFL activities based on authentic texts (and a range of other uses, like readability assessment, test format preparation, lexical frequency profile, adapting texts to international readership, etc.)	
Platform:	PC	
Minimum hardware requirements:	OS Version: Microsoft Windows 1998, 2000, ME, XP.	
Authors:	Program:	Jaak Denies
	Technical advice:	J. Ceulemans, M. Decuyper, A. Van Gompel
	Educational advice:	Michaël Goethals, (K.U.Leuven), Karel Van Rompaey, Gilbert Deketelaere, Marc Hoefkens
	User's Manual:	Michaël Goethals Faculty of Arts, EFL Teacher training Unit, Blijde-Inkomststraat 21, B-3000 Leuven, Belgium tel. +32 16 32 47 64 fax +32 16 32 55 47 michael.goethals@arts.kuleuven.ac.be
Price:	Free. Downloadable from: http://engels.vvks-ict.com/engict/wordclassifier2004.zip ; 12.5 Mb zip file includes user's manual (in pdf)	

Editor's note: The author of the following review was involved in the creation of the program. However, since the program is available to download for free, it seems clear that the review's purpose is to inform the TESL-EJ readership, and not to further any commercial interest. Therefore, it was decided that the review is a worthwhile contribution to TESL-EJ.

General description

WordClassifier classifies words of a text into lists, according to learning ranges in a sequence of their probable usefulness for learners of EILE (English as an International Language in Europe) as listed in the E.E.T-list (the *European English Teaching Vocabulary-list*, prepared and updated by the EFL Teacher Training Unit of the Faculty of Arts of the K.U.Leuven)

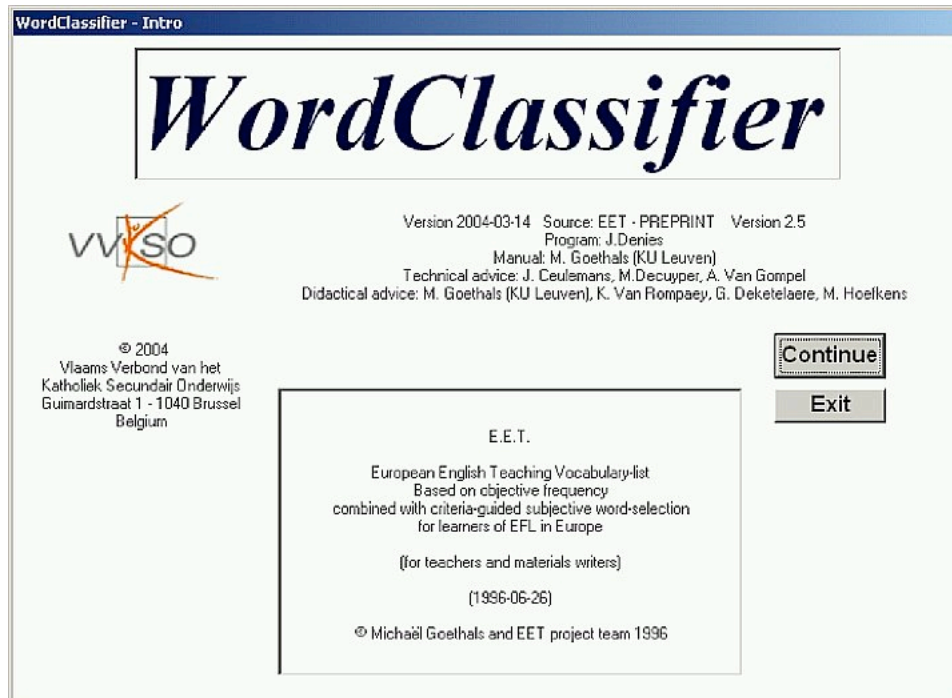


Figure 1—Opening screen

Basically, *WordClassifier* reshuffles the words of a text into lists of words according to a number of criteria, of which frequency is the most prominent. The classification is figured per "word cluster" (or "word family": Headword and semantically problem-free derivations).

The user can copy text from any text file, (e.g., a MS-Word document or passages selected from the Internet) and paste it into *WordClassifier's* input window for word classification analysis. See Figure 2.

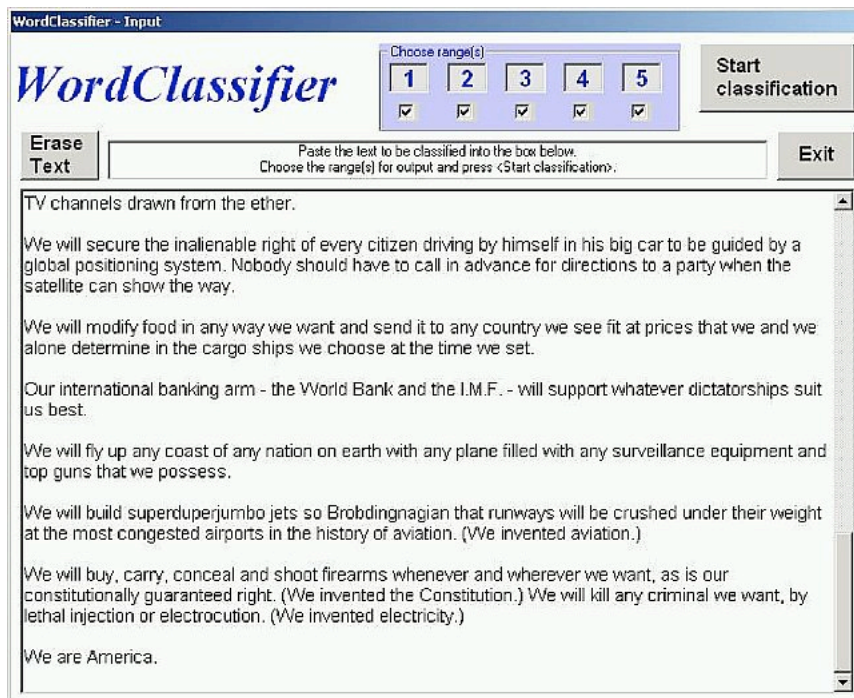


Figure 2—Text input screen, filled with a newspaper article from the Internet

WordClassifier classifies the words in a text into seven different lists:

- Range 0: Grammatical words
- Range 1: Complete beginners' vocabulary
- Range 2: More words for the lower level learner
- Range 3, range 4, and range 5: Words belonging to the next three ranges of the next most useful group of words
- Outsiders: The remaining words that WordClassifier does not recognize as being in the program's word-list for the learning ranges one through five

Each list is presented in a box (see Figure 3), from which it can then be selected and copied for further use, or copied and saved as a separate text file. Buttons allow the user to save the output as a report (see Figure 4) or to give the words of the text a different color according to their learning range (see Figure 5).

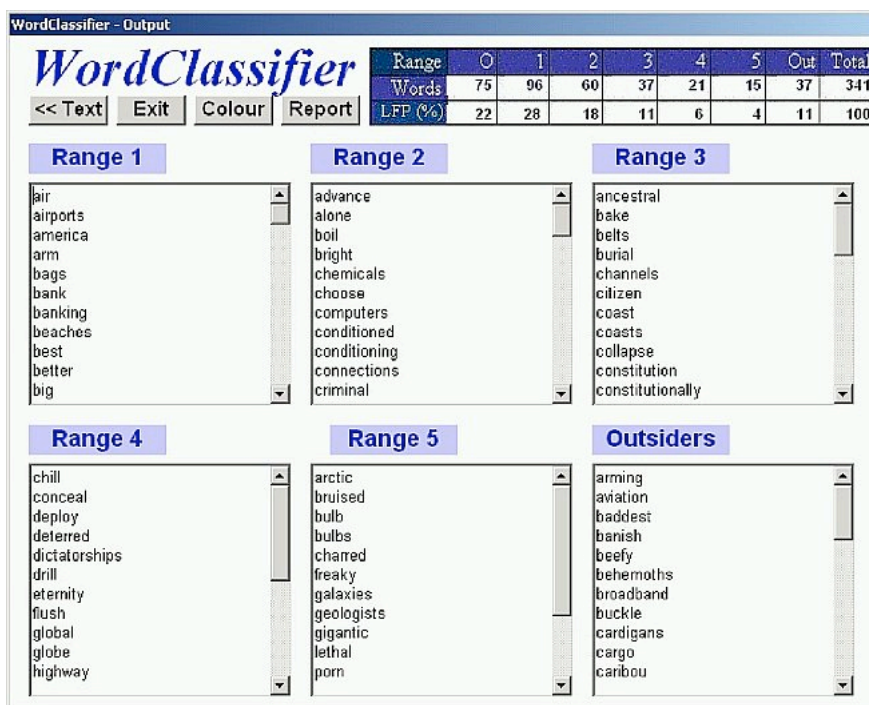


Figure 3—Output screen: the Lexical Frequency Profile (LFP) and the words in 6 boxes (after pruning away proper names, abbreviations, etc.)

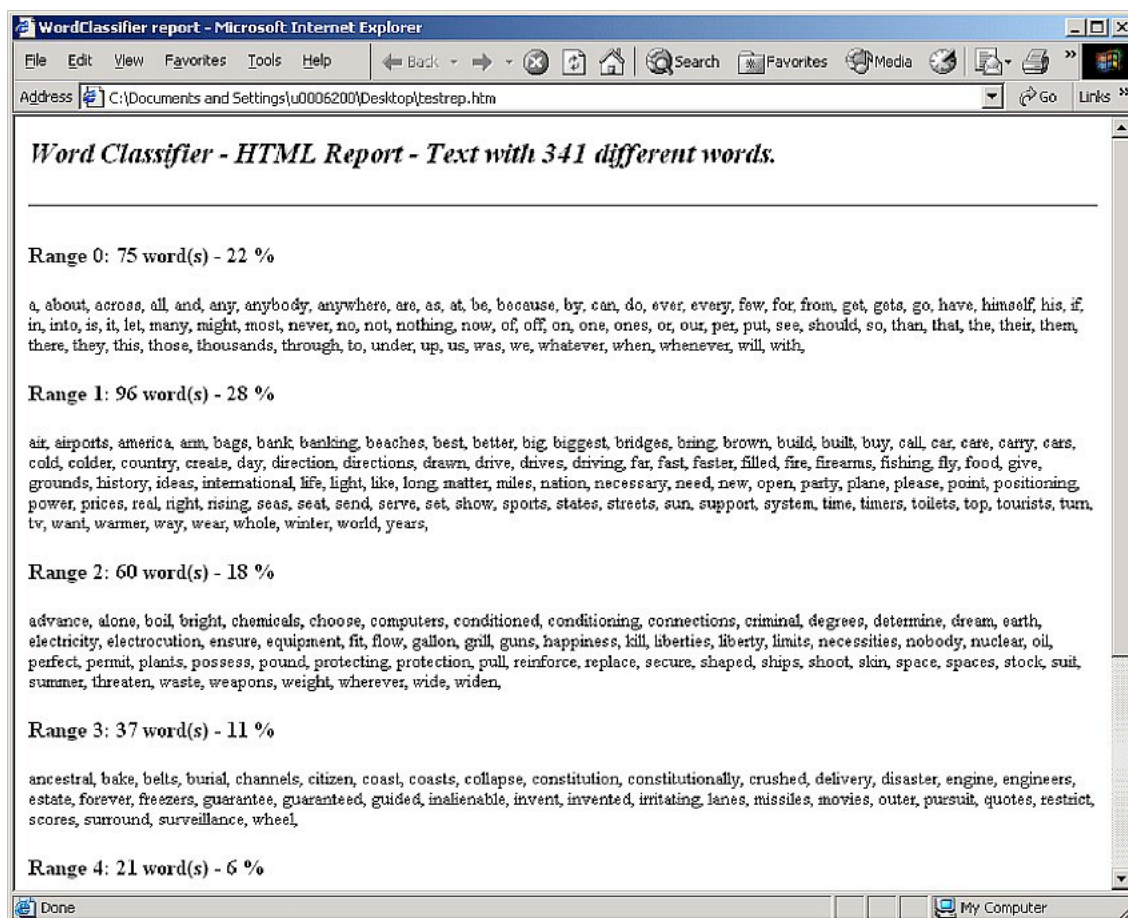


Figure 4—Report in list format

The user can choose which boxes are to be filled.

With the subsets of words comes a *lexical frequency profile* (LFP) (see Figure 5), presenting the number and percentages of different word forms ('types', not 'tokens', for those initiated in corpus linguistics and statistics) in each subset. As a whole, this provides a good, direct image of the readability of a text, although the difficulty of the content, argument, and style will still need to be taken into account.

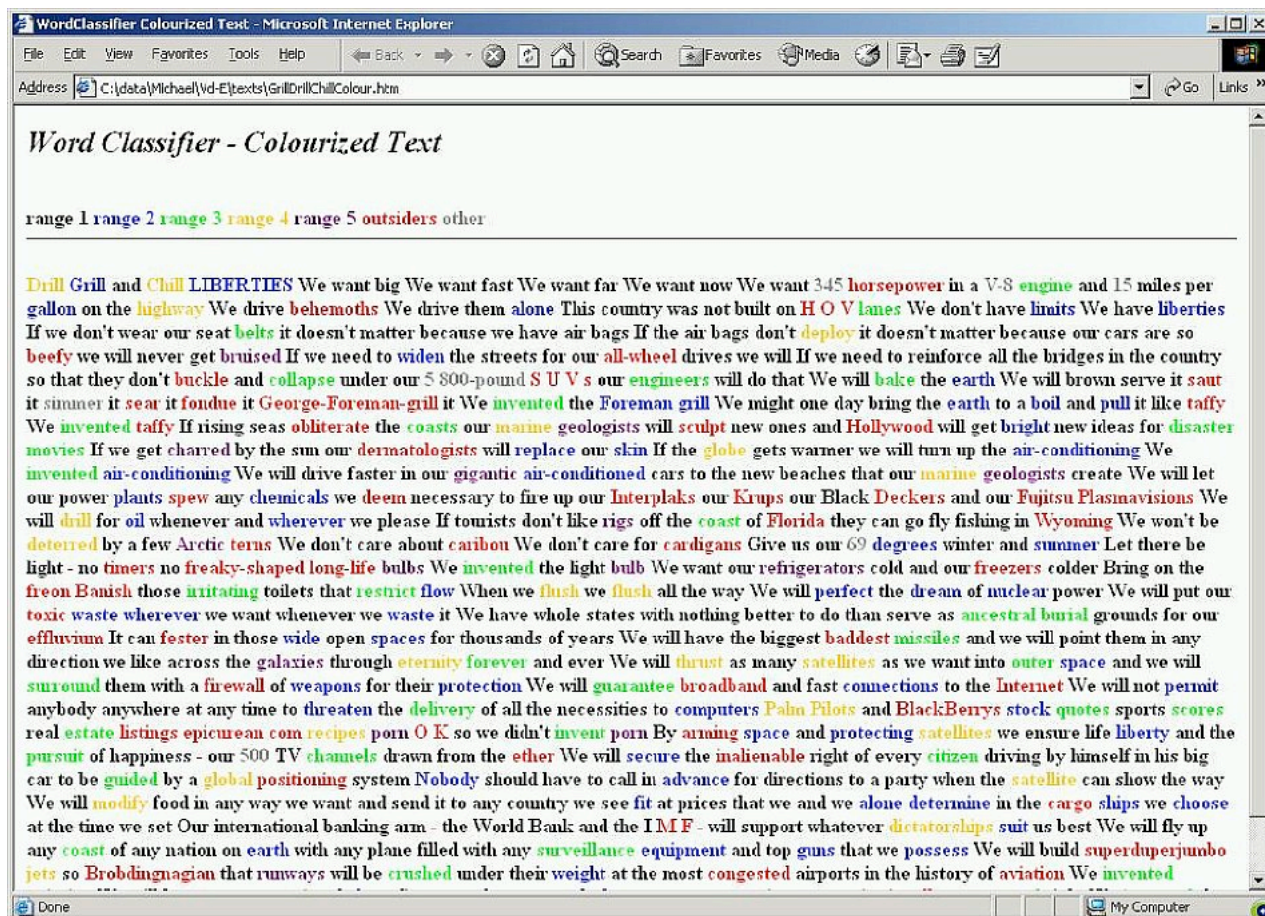


Figure 5—Report in colours per range

As was mentioned, the user can generate a report file with the words in separate lists (see figure 4), or a report with all the word forms in the order of the text but colored according to the subsets. Such reports appear without punctuation, but a materials designer can re-introduce punctuation manually if desired. This color report allows direct access to the difficult words and passages of a text while at the same time highlighting typos, proper names, abbreviations, and the like. Such features can be corrected, adapted, or deleted, so that a "pruned" text can be copied back into the text screen for a second analysis. Such a second analysis should provide an enhanced LFP and a more directly usable set of lists.

The pedagogical idea behind the program is the concept of vocabulary management, or mastering the chaotic offerings of "difficult" words in authentic texts. In this method, the text is not edited by ESOL authors or teachers, and so is not artificially simplified. For words in the earlier, perhaps already acquired levels it offers suggestions for remedial vocabulary work (i.e., "This is a frequently used word that you could be supposed to know already"). For the more difficult levels it allows users to make a reasonable choice regarding which words are important to learn and practice first, and which words can be left for later study, or for incidental learning according to interest or professional need.

Evaluation

WordClassifier allows ESOL materials writers and teachers using authentic texts for classroom activities to either simplify texts to levels appropriate for their learners, or (as the authors certainly favor) to steer learners' limited time and energy to learning what is most useful. This should lead to maximum learning gains for immediate, daily use and provide an optimal basis for further learning by highlighting the next set of words that are likely to be of most use to the learner group or individual. The words beyond that set might be de-emphasized or not focused on in learning situations. This makes the use of authentic materials more manageable in communicative, natural methodological settings.

The alphabetical word lists generated by *WordClassifier* are ready to be combined with dictionary software (for the preparation of explanatory word-lists) or exercise generating authorware (in their simplest form gap-filling exercises).

Of course, the speed of the program includes some inherent limitations. This program is not a parser (with all the ambiguity problems that would even then remain if it were one). It merely compares the word forms in a text to the

classification values listed in the database behind the program, the *EET-vocabulary list*. It cannot handle figurative language (metaphor), multi-word units, transparent words, or false friends. Identifying those is the teacher's task, and this is where the teacher's professionalism is most needed, although informed, autonomous, and probably advanced learners might also be able to make sound judgments using the program.

The program is certainly a time-saver compared to using dictionaries and frequency lists and consulting colleagues about the selection of words for learning. The teacher only has to take decisions about the 2 to 5 % of words or collocations that are problematic and decide whether or not to include them as targets for study. For learners, it is a liberation from having to study *all* the new words in a text.

For the preparation of gap-filling exercises and crossword puzzles in ESOL contexts *WordClassifier* application is self-evident. It also seems to offer new perspectives for literary analysis as well, especially in ESOL contexts.

Installation is simple and straightforward. In fact, the program is an MS Access application with an encrypted set of data (so users neither have to worry about, nor can they change the data of the word-list). The installation leaves a shortcut for immediate access to the tool.

The LFP is a measure that can assist writers of texts for an international, non-native speaker readership. Some members of the BBC English staff have reported using *WordClassifier* occasionally for this purpose (personal e-mail communication). For use in general tests of learner growth (e.g., as part of a language portfolio) the program has represented that growth very consistently. For advanced EFL students in higher education, the LFP has proved to be a valuable measure of learner growth.

Colleagues teaching EAP courses such as business English have reported training their students to use *WordClassifier* to plan their own learning of professional vocabulary and in preparing presentations and activities for peer-learning (Baten & Goethals 2005).

Finally, a word about the "outsiders" range is perhaps in order. The fact that the database is the result of systematic clustering of actually occurring words in a set of text corpora (see Goethals 1992, 2001, and 2004—the last article gives a good survey of the design principles of the *E.E.T. Vocabulary-list*), means that a good number of words that western European EFL learners may be familiar with as loanwords in their own language, but that happen not to have appeared in the texts that were collected for the text corpora, are classified as outsiders. This sometimes comes as a surprise to learners and teachers at first sight. The same goes for new or fashionable words that have come into frequent and familiar use since the time those corpora were collected. The same also goes for many derivations (plurals, genitive forms, -ly forms, -ing forms and -ed forms) that happened not to be in those corpora. A good number of those have been added manually, but this approach is practically impossible. Simple compound words, like headache, formed by two high frequency words, may be relatively infrequent as compounds in the corpus, but most such forms that were actually in the texts have been reclassified into the ranges of the less frequent of the two clusters to which they belong. There are quite a number of other such technicalities that cannot be discussed here. As was mentioned above, this is where teachers or learner-users have to judge for themselves, keeping in mind that this program has no ambition to replace the teacher or the autonomous learner as the real decision taker in learning matters: It only wants to assist those decisions.

References

- Baten, L. & Goethals, M. (2005). Lexical Frequency as a criterion and a tool in teaching and learning ESP. (Forthcoming, can be requested from the author of this review.)
- Goethals, M. (1992). COBUILD, BNC, LET, LCL, Marzano and the others. Forging an instrument for vocabulary learning/teaching from word frequency counts, word clusters and other types of vocabulary lists. *ITL Review of Applied Linguistics* 97-98, 121-158.
- Goethals, M. (1994). Vocabulary management in foreign language teaching and learning: The concept of vocabulary management and a case study into EFL teachers' selection of target vocabulary from an authentic text. In K. Carlon, K. Davidse & B. Rudzka-Ostyn (Eds.), *Perspectives on English: Studies in honor of Prof. Dr. Emma Vorlat* (pp. 484-506). Leuven: Universitaire Pers.
- Goethals, M. (1996). *E.E.T., European English Teaching Vocabulary-list, based on objective frequency combined with criteria-guided subjective word-selection for learners of EFL in Europe*. Version 1.0-b. Leuven: K.U.Leuven (Faculty of Arts, Teacher Training Unit).
- Goethals, M. (2001). The use of word frequency data in the teaching of English as an alternative/additional language: Reflections on recent EET-list experience and experiments. In M. Bax, & J.W. Zwart (Eds.), *Reflections on Language and Language Learning. In honor of Arthur van Essen* (pp. 311-323). Amsterdam/Philadelphia: John Benjamins.
- Goethals, M. (2004). E.E.T.: The European English Teaching Vocabulary-list. A presentation of the project concepts and procedures. In B. Lewandowska-Tomaszczyk (Ed.), *Practical Applications in Language and Computers* (PALC 2003) (pp. 417-427). Hamburg: Peter Lang.

About the Reviewer

Michaël Goethals is an applied linguist, EFL teacher trainer, and in-service teaching organizer with more than 25 years experience. His main area of research is in using lexical frequency for ESOL. He has also worked as a curriculum planner and coordinator of curricula for modern languages in the Flemish (Belgian) Catholic school network. In addition, he has served as coordinator of the EU Lingua, Comenius and Socrates-Minerva projects. A former guest lecturer in The Netherlands, China, Vietnam, Indonesia and The Philippines, he was also organizer of Euro CALL 1998 in Leuven, Belgium, and is a member of the Editorial Board of ReCALL.

© Copyright rests with authors. Please cite TESL-EJ appropriately.

Editor's Note: The HTML version contains no page numbers. Please use the [PDF version](#) of this article for citations.