Appendixes – *ReaderBench* Workflows, Print-Screens and Input Examples

As it can be observed from Figure 62 and Figure 63, *ReaderBench* consists of three main tabs, each focused on specific functionalities: 1/ *document* assessment, both for general texts and conversations, 2/ *verbalization* assessment and 3/ *textual complexity* corpus training and evaluation. The conventions of all the workflows are the following: *arcs* denote possible transition, whereas *rectangles* signify actions (flat bottom rectangle) or other interfaces depicted as figures within the book (wavy bottom rectangle).

**Fig. 62** *ReaderBench* General workflow

**Fig. 63** *ReaderBench* Main user interface
Fig. 64 ReaderBench Document workflow
Fig. 65 ReaderBench Document management interface. Enables the possibilities for the tutor to create new, load, edit and save texts (reading materials) in corresponding XML format, with all recommended fields, including the possibility to define verbalization breakpoints that are automatically considered when generating new learner self-explanations (see Figure 71).

Fig. 66 ReaderBench Document processing interface. Allows users to add, remove or visualize a loaded document (conversation or general text). Serialized documents are pre-computed documents that are saved as serialized Java objects and can be easily recovered in order to eliminate the processing time required for a new document.
Fig. 67 *ReaderBench* Interface for adding a new document for processing

Fig. 68 *ReaderBench* Document advanced visualization. An advanced view available only for documents that presents cohesion-based scores using different semantic measures for adjacent sentences of the same paragraph.
Fig. 69 *ReaderBench* Voice selection interface. This interface enables the user to manually select the voices (semantic word chains) of interest, ordered in descending order of the number of comprised concepts, later to be used for displaying voice inter-animation (at least one voice must be selected) (see Figure 38). By default, as the number of overall voices can become cumbersome, all voices are deselected. A voice is displayed as a tuple of the 3 most frequently occurring word lemmas within the semantic chain. The lower part of the interface displays the entire selected semantic chain, consisting of flectional word forms followed by paragraph ID/sentence ID (general texts) or utterance ID/sentence ID (conversations or forum discussion threads).

**Appendix B – Verbalization Workflow and Additional Print-Screens**

![Verbalization workflow diagram](image)

Fig. 70 *ReaderBench* Verbalization workflow
Fig. 71 ReaderBench Interface for creating new self-explanations. An interface designed for learners, enabling them to add their self-explanations in an intuitive manner, after each corresponding chunk of text from the original reading material.

Fig. 72 ReaderBench Interface for manually annotating self-explanations. Tutors are granted the possibility to load learner self-explanations and annotate them with the correspondingly identified reading strategies.
Fig. 73 ReaderBench Verbalization processing interface Add, view and deleted operations on learner verbalizations. At least one document must be loaded.

Fig. 74 ReaderBench Interface for adding a new verbalization for processing
Appendix C – Textual Complexity Additional Print-Screen and EA/AA Scores

**Fig. 75** *ReaderBench* Corpus textual complexity assessment interface. Determines all textual complexity factors for all the documents within the training corpora. All the documents for a specific complexity class are included as separate XML files within a folder named “class<ID>“. In addition, the user can select the LSA and LDA semantic models he/she wants to use in the assessment process and whether part-of-speech tagging should be applied on the corpus or not.

The detailed evaluation results in terms of all complexity dimensions and all withheld textual complexity factors are presented in Table 40. The factors with a high agreement (EA ≥ 0.5 and AA ≥ .85) are marked with *italics* and “*“.

**Table 40** Exact Agreement (EA) and Adjacent Agreement (AA) for all evaluation factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>C1(%)</th>
<th>C2(%)</th>
<th>C3(%)</th>
<th>C4(%)</th>
<th>C5(%)</th>
<th>C6(%)</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
</tr>
<tr>
<td>All Factors Combined*</td>
<td>.9/1</td>
<td>.8/1</td>
<td>.76/1</td>
<td>.7/99</td>
<td>.7/1</td>
<td>.82/99</td>
<td>.78/1</td>
</tr>
<tr>
<td>Textual Complexity Dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Readability Factors*</td>
<td>.82/1</td>
<td>.72/1</td>
<td>.68/1</td>
<td>.57/99</td>
<td>.68/99</td>
<td>.78/98</td>
<td>.71/99</td>
</tr>
<tr>
<td>Fluency Factors</td>
<td>.74/76</td>
<td>.03/57</td>
<td>.18/28</td>
<td>.14/37</td>
<td>.19/68</td>
<td>.64/76</td>
<td>.32/57</td>
</tr>
<tr>
<td>Structure Complexity Factors*</td>
<td>.91/1</td>
<td>.76/99</td>
<td>.61/1</td>
<td>.62/97</td>
<td>.66/99</td>
<td>.73/99</td>
<td>.72/99</td>
</tr>
<tr>
<td>Diction Factors*</td>
<td>.85/1</td>
<td>.61/97</td>
<td>.43/9</td>
<td>.41/86</td>
<td>.26/86</td>
<td>.71/86</td>
<td>.55/91</td>
</tr>
<tr>
<td>Entropy Factors</td>
<td>.58/68</td>
<td>.23/62</td>
<td>.07/3</td>
<td>.12/35</td>
<td>.2/72</td>
<td>.56/72</td>
<td>.3/56</td>
</tr>
<tr>
<td>Balanced CAF Factors*</td>
<td>.87/1</td>
<td>.81/1</td>
<td>.72/1</td>
<td>.68/99</td>
<td>.66/1</td>
<td>.77/99</td>
<td>.75/1</td>
</tr>
<tr>
<td>Part of Speech Complexity Factors*</td>
<td>.87/99</td>
<td>.58/98</td>
<td>.5/95</td>
<td>.35/91</td>
<td>.41/89</td>
<td>.67/87</td>
<td>.56/93</td>
</tr>
<tr>
<td>Parsing Tree Complexity Factors</td>
<td>.73/96</td>
<td>.45/88</td>
<td>.18/74</td>
<td>.35/62</td>
<td>.13/8</td>
<td>.65/76</td>
<td>.42/79</td>
</tr>
<tr>
<td>Named Entity Complexity Factors*</td>
<td>.87/1</td>
<td>.59/96</td>
<td>.49/92</td>
<td>.44/88</td>
<td>.37/9</td>
<td>.69/87</td>
<td>.58/92</td>
</tr>
<tr>
<td>Co-reference Complexity Factors</td>
<td>.68/9</td>
<td>.34/78</td>
<td>.26/64</td>
<td>.29/62</td>
<td>.17/79</td>
<td>.45/71</td>
<td>.37/74</td>
</tr>
</tbody>
</table>
Table 40 (continued)

<table>
<thead>
<tr>
<th>Factor</th>
<th>C1(%)</th>
<th>C2(%)</th>
<th>C3(%)</th>
<th>C4(%)</th>
<th>C5(%)</th>
<th>C6(%)</th>
<th>Avg. (EA/A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
</tr>
<tr>
<td><strong>Word Complexity Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.71/.98</td>
<td>.52/.98</td>
<td>.47/.87</td>
<td>.34/.88</td>
<td>.52/.9</td>
<td>.73/.95</td>
<td>.55/.93</td>
</tr>
<tr>
<td><strong>Lexical Chains Factors</strong></td>
<td>.62/.85</td>
<td>.41/.8</td>
<td>.21/.62</td>
<td>.23/.54</td>
<td>.13/.75</td>
<td>.58/.72</td>
<td>.36/.71</td>
</tr>
<tr>
<td><strong>Discourse Factors</strong></td>
<td>.78/.98</td>
<td>.49/.94</td>
<td>.42/.86</td>
<td>.38/.82</td>
<td>.39/.91</td>
<td>.69/.84</td>
<td>.53/.89</td>
</tr>
<tr>
<td><strong>Readability Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Readability FOG</strong></td>
<td>.72/.95</td>
<td>.52/.94</td>
<td>.28/.87</td>
<td>.34/.84</td>
<td>.49/.94</td>
<td>.74/.95</td>
<td>.51/.91</td>
</tr>
<tr>
<td><strong>Readability Kincaid</strong></td>
<td>.72/1</td>
<td>.57/.99</td>
<td>.53/.95</td>
<td>.39/.95</td>
<td>.54/.97</td>
<td>.79/.98</td>
<td>.59/.97</td>
</tr>
<tr>
<td><strong>Number of words per sentence</strong></td>
<td>.76/.98</td>
<td>.62/.95</td>
<td>.4/.93</td>
<td>.38/.88</td>
<td>.38/.91</td>
<td>.68/.86</td>
<td>.54/.92</td>
</tr>
<tr>
<td><strong>Average number of syllables per word</strong></td>
<td>.73/.97</td>
<td>.46/.98</td>
<td>.52/.83</td>
<td>.27/.83</td>
<td>.47/.87</td>
<td>.72/.98</td>
<td>.53/.91</td>
</tr>
<tr>
<td><strong>Percentage of complex words (&gt;2 syllables)</strong></td>
<td>.71/.92</td>
<td>.32/.91</td>
<td>.28/.69</td>
<td>.17/.71</td>
<td>.35/.8</td>
<td>.72/.95</td>
<td>.42/.83</td>
</tr>
<tr>
<td><strong>Fluency Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Normalized number of words</strong></td>
<td>.72/.75</td>
<td>.06/.58</td>
<td>.27/.31</td>
<td>.03/.23</td>
<td>.01/.54</td>
<td>.67/.73</td>
<td>.29/.52</td>
</tr>
<tr>
<td><strong>Structure Complexity Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Normalized number of blocks</strong></td>
<td>.79/.95</td>
<td>.3/.83</td>
<td>.05/.23</td>
<td>.07/.19</td>
<td>.12/.74</td>
<td>.78/.87</td>
<td>.35/.63</td>
</tr>
<tr>
<td><strong>Average block size</strong></td>
<td>.76/.97</td>
<td>.48/.9</td>
<td>.3/.68</td>
<td>.18/.45</td>
<td>.26/.84</td>
<td>.68/.77</td>
<td>.44/.77</td>
</tr>
<tr>
<td><strong>Normalized number of sentences</strong></td>
<td>.7/.97</td>
<td>.51/.85</td>
<td>.15/.77</td>
<td>.41/.57</td>
<td>.08/.86</td>
<td>.78/.81</td>
<td>.44/.8</td>
</tr>
<tr>
<td><strong>Average sentence length</strong></td>
<td>.82/1</td>
<td>.6/.97</td>
<td>.41/.86</td>
<td>.4/.84</td>
<td>.29/.86</td>
<td>.67/.86</td>
<td>.53/.9</td>
</tr>
<tr>
<td><strong>Diction Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average word length</strong></td>
<td>.84/1</td>
<td>.63/.98</td>
<td>.44/.91</td>
<td>.41/.89</td>
<td>.26/.88</td>
<td>.66/.86</td>
<td>.54/.92</td>
</tr>
<tr>
<td><strong>Entropy Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Word entropy</strong></td>
<td>.68/.88</td>
<td>.44/.91</td>
<td>.02/.38</td>
<td>.02/.08</td>
<td>.05/.18</td>
<td>.11/.16</td>
<td>.22/.43</td>
</tr>
<tr>
<td><strong>Character entropy</strong></td>
<td>.56/.56</td>
<td>.04/.53</td>
<td>.1/.28</td>
<td>.24/.32</td>
<td>.0/.78</td>
<td>.71/.71</td>
<td>.27/.53</td>
</tr>
<tr>
<td><strong>Balanced CAF Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lexical Diversity</strong></td>
<td>.77/.88</td>
<td>.22/.87</td>
<td>.33/.46</td>
<td>.09/.37</td>
<td>.05/.38</td>
<td>.26/.35</td>
<td>.29/.55</td>
</tr>
<tr>
<td><strong>Lexical Sophistication</strong></td>
<td>.84/1</td>
<td>.78/1</td>
<td>.69/1</td>
<td>.59/.99</td>
<td>.58/1</td>
<td>.81/.98</td>
<td>.71/1</td>
</tr>
<tr>
<td><strong>Syntactic Diversity</strong></td>
<td>.71/.74</td>
<td>.11/.77</td>
<td>.22/.57</td>
<td>.25/.62</td>
<td>.14/.38</td>
<td>0/.11</td>
<td>.24/.53</td>
</tr>
<tr>
<td><strong>Syntactic Sophistication</strong></td>
<td>.85/.99</td>
<td>.57/.98</td>
<td>.49/.95</td>
<td>.34/.87</td>
<td>.36/.88</td>
<td>.7/.85</td>
<td>.55/.92</td>
</tr>
<tr>
<td><strong>Balanced CAF</strong></td>
<td>.83/.99</td>
<td>.61/.99</td>
<td>.54/.84</td>
<td>.16/.71</td>
<td>.24/.62</td>
<td>.5/.68</td>
<td>.48/.8</td>
</tr>
</tbody>
</table>
Table 40 (continued)

<table>
<thead>
<tr>
<th>Factor</th>
<th>C1(%) EA/A</th>
<th>C2(%) EA/A</th>
<th>C3(%) EA/A</th>
<th>C4(%) EA/A</th>
<th>C5(%) EA/A</th>
<th>C6(%) EA/A</th>
<th>Avg. EA/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part of Speech Complexity Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of nouns*</td>
<td>.86/1</td>
<td>.57/.96</td>
<td>.5/.9</td>
<td>.35/.81</td>
<td>.23/.86</td>
<td>.66/.82</td>
<td>.53/.89</td>
</tr>
<tr>
<td>Average number of pronouns</td>
<td>.77/.89</td>
<td>.17/.77</td>
<td>.08/.29</td>
<td>.08/.19</td>
<td>.05/.53</td>
<td>.45/.51</td>
<td>.27/.53</td>
</tr>
<tr>
<td>Average number of verbs</td>
<td>.75/.89</td>
<td>.21/.71</td>
<td>.09/.4</td>
<td>.13/.31</td>
<td>.15/.75</td>
<td>.67/.76</td>
<td>.33/.64</td>
</tr>
<tr>
<td>Average number of adverbs</td>
<td>.63/.89</td>
<td>.3/.88</td>
<td>.16/.36</td>
<td>0/.21</td>
<td>.16/.29</td>
<td>.25/.34</td>
<td>.25/.5</td>
</tr>
<tr>
<td>Average number of adjectives</td>
<td>.77/.95</td>
<td>.44/.95</td>
<td>.38/.8</td>
<td>.31/.75</td>
<td>.22/.72</td>
<td>.43/.65</td>
<td>.42/.8</td>
</tr>
<tr>
<td>Average number of prepositions</td>
<td>.76/.99</td>
<td>.59/.93</td>
<td>.31/.85</td>
<td>.4/.71</td>
<td>.07/.82</td>
<td>.65/.7</td>
<td>.46/.83</td>
</tr>
<tr>
<td>Parsing Tree Complexity Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average tree depth</td>
<td>.65/.93</td>
<td>.39/.81</td>
<td>.2/.73</td>
<td>.36/.6</td>
<td>.1/.79</td>
<td>.66/.71</td>
<td>.39/.76</td>
</tr>
<tr>
<td>Average tree size</td>
<td>.75/.97</td>
<td>.45/.86</td>
<td>.18/.74</td>
<td>.39/.57</td>
<td>.08/.83</td>
<td>.74/.8</td>
<td>.43/.79</td>
</tr>
<tr>
<td>Named Entity Complexity Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of named entities</td>
<td>.63/.76</td>
<td>.17/.73</td>
<td>.12/.26</td>
<td>.01/.07</td>
<td>.03/.5</td>
<td>.69/.71</td>
<td>.28/.5</td>
</tr>
<tr>
<td>Total number of entities per document</td>
<td>.67/.88</td>
<td>.33/.68</td>
<td>0/.38</td>
<td>.26/.37</td>
<td>.24/.69</td>
<td>.37/.53</td>
<td>.31/.59</td>
</tr>
<tr>
<td>Total number of unique entities per document</td>
<td>.61/.88</td>
<td>.27/.63</td>
<td>.03/.44</td>
<td>.29/.34</td>
<td>.06/.65</td>
<td>.49/.52</td>
<td>.29/.58</td>
</tr>
<tr>
<td>Percentage of entities per document</td>
<td>.73/.76</td>
<td>.05/.61</td>
<td>.09/.29</td>
<td>.2/.44</td>
<td>.18/.7</td>
<td>.44/.56</td>
<td>.28/.56</td>
</tr>
<tr>
<td>Percentage of unique entities per document</td>
<td>.48/.78</td>
<td>.37/.61</td>
<td>.11/.51</td>
<td>.24/.41</td>
<td>.1/.65</td>
<td>.34/.42</td>
<td>.27/.56</td>
</tr>
<tr>
<td>Average number of entities per sentence*</td>
<td>.85/1</td>
<td>.54/.98</td>
<td>.44/.91</td>
<td>.34/.79</td>
<td>.22/.83</td>
<td>.67/.82</td>
<td>.51/.89</td>
</tr>
<tr>
<td>Average number of unique entities per sentences</td>
<td>.87/1</td>
<td>.48/.97</td>
<td>.35/.77</td>
<td>.2/.6</td>
<td>.19/.64</td>
<td>.45/.55</td>
<td>.42/.75</td>
</tr>
<tr>
<td>Percentage of named entities per document</td>
<td>.63/.73</td>
<td>.11/.7</td>
<td>.08/.16</td>
<td>.05/.07</td>
<td>.04/.56</td>
<td>.72/.74</td>
<td>.27/.49</td>
</tr>
<tr>
<td>Average number of named entities per sentence</td>
<td>.74/.87</td>
<td>.24/.79</td>
<td>.19/.46</td>
<td>.06/.29</td>
<td>.13/.41</td>
<td>.34/.36</td>
<td>.28/.53</td>
</tr>
<tr>
<td>Percentage of named entities in total entities</td>
<td>.67/.75</td>
<td>.11/.72</td>
<td>.06/.2</td>
<td>.05/.16</td>
<td>.08/.62</td>
<td>.76/.81</td>
<td>.29/.54</td>
</tr>
<tr>
<td>Percentage of nouns in total entities</td>
<td>.73/.84</td>
<td>.09/.78</td>
<td>.05/.22</td>
<td>.04/.15</td>
<td>.12/.59</td>
<td>.68/.74</td>
<td>.28/.55</td>
</tr>
<tr>
<td>Percentage of nouns per document</td>
<td>.48/.73</td>
<td>.37/.67</td>
<td>.08/.54</td>
<td>.21/.44</td>
<td>.16/.79</td>
<td>.69/.83</td>
<td>.33/.67</td>
</tr>
<tr>
<td>Average number of nouns per</td>
<td>.83/.98</td>
<td>.59/.94</td>
<td>.53/.95</td>
<td>.44/.89</td>
<td>.35/.91</td>
<td>.65/.89</td>
<td>.57/.93</td>
</tr>
</tbody>
</table>
Table 40 (continued)

<table>
<thead>
<tr>
<th>Factor</th>
<th>C1(%)</th>
<th>C2(%)</th>
<th>C3(%)</th>
<th>C4(%)</th>
<th>C5(%)</th>
<th>C6(%)</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
<td>EA/A</td>
</tr>
<tr>
<td><strong>sentence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage remaining nouns per document</td>
<td>.55/8</td>
<td>.3/.77</td>
<td>.21/.57</td>
<td>.21/.34</td>
<td>0/.66</td>
<td>.63/.63</td>
<td>.32/.63</td>
</tr>
<tr>
<td>Average number of remaining nouns per sentence</td>
<td>.85/1</td>
<td>.54/.97</td>
<td>.48/.84</td>
<td>.23/.68</td>
<td>.29/.77</td>
<td>.58/.8</td>
<td>.5/84</td>
</tr>
<tr>
<td>Percentage of overlapping nouns per document</td>
<td>.48/.76</td>
<td>.22/.67</td>
<td>.03/.24</td>
<td>.01/.02</td>
<td>0/.6</td>
<td>.78/.8</td>
<td>.25/.52</td>
</tr>
<tr>
<td>Average number of overlapping nouns per sentence</td>
<td>.82/.85</td>
<td>.12/.79</td>
<td>.17/.33</td>
<td>.06/.23</td>
<td>.1/.55</td>
<td>.44/.5</td>
<td>.28/.54</td>
</tr>
<tr>
<td><strong>Co-reference Complexity Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of co-reference chains per document</td>
<td>.14/.41</td>
<td>.26/.55</td>
<td>.13/.47</td>
<td>.19/.44</td>
<td>.21/.72</td>
<td>.38/.59</td>
<td>.22/.53</td>
</tr>
<tr>
<td>Average number of co-references per chain</td>
<td>.42/.47</td>
<td>.11/.43</td>
<td>.16/.32</td>
<td>.1/.31</td>
<td>.08/.8</td>
<td>.76/.82</td>
<td>.27/.52</td>
</tr>
<tr>
<td>Average co-reference chain span</td>
<td>.65/.82</td>
<td>.27/.65</td>
<td>.01/.41</td>
<td>.33/.36</td>
<td>.04/.64</td>
<td>.38/.41</td>
<td>.28/.55</td>
</tr>
<tr>
<td>Number of co-reference chains with a big span</td>
<td>.15/.47</td>
<td>.33/.58</td>
<td>.16/.4</td>
<td>.06/.38</td>
<td>.21/.68</td>
<td>.53/.72</td>
<td>.24/.54</td>
</tr>
<tr>
<td>Average inference distance per co-reference chain</td>
<td>.63/.79</td>
<td>.31/.84</td>
<td>.1/.34</td>
<td>0/.2</td>
<td>.24/.41</td>
<td>.16/.36</td>
<td>.24/.49</td>
</tr>
<tr>
<td>Number of active co-reference chains per word</td>
<td>.29/.53</td>
<td>.33/.64</td>
<td>.19/.57</td>
<td>.23/.5</td>
<td>.11/.62</td>
<td>.4/.5</td>
<td>.26/.56</td>
</tr>
<tr>
<td>Number of active co-reference chains per entity</td>
<td>.4/.68</td>
<td>.29/.82</td>
<td>.22/.54</td>
<td>.05/.27</td>
<td>.02/.36</td>
<td>.41/.45</td>
<td>.23/.52</td>
</tr>
<tr>
<td><strong>Word Complexity Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean distance between lemma and word stems</td>
<td>.6/.9</td>
<td>.38/.93</td>
<td>.3/.76</td>
<td>.32/.74</td>
<td>.43/.83</td>
<td>.62/.88</td>
<td>.44/.84</td>
</tr>
<tr>
<td>Mean distance between words and corresponding stems</td>
<td>.61/.85</td>
<td>.25/.91</td>
<td>.32/.72</td>
<td>.27/.79</td>
<td>.48/.82</td>
<td>.62/.9</td>
<td>.42/.83</td>
</tr>
<tr>
<td>Mean word distance in hypernym tree</td>
<td>.09/.31</td>
<td>.2/.67</td>
<td>.52/.73</td>
<td>.08/.7</td>
<td>.15/.35</td>
<td>.1/.25</td>
<td>.19/.5</td>
</tr>
<tr>
<td>Mean word polysemy count*</td>
<td>.73/.97</td>
<td>.5/.98</td>
<td>.54/.89</td>
<td>.43/.86</td>
<td>.45/.9</td>
<td>.75/.97</td>
<td>.57/.93</td>
</tr>
<tr>
<td>Mean word syllable count*</td>
<td>.76/.98</td>
<td>.51/.98</td>
<td>.56/.89</td>
<td>.35/.84</td>
<td>.47/.88</td>
<td>.72/.96</td>
<td>.56/.92</td>
</tr>
<tr>
<td><strong>Lexical Chains Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average span of lexical chains</td>
<td>.56/.77</td>
<td>.29/.78</td>
<td>.2/.45</td>
<td>.05/.37</td>
<td>.09/.27</td>
<td>.1/.18</td>
<td>.22/.47</td>
</tr>
</tbody>
</table>
### Table 40 (continued)

<table>
<thead>
<tr>
<th>Factor</th>
<th>C1(%)</th>
<th>C2(%)</th>
<th>C3(%)</th>
<th>C4(%)</th>
<th>C5(%)</th>
<th>C6(%)</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lexical chains with more than 5 concepts</td>
<td>.66/.89</td>
<td>.22/.69</td>
<td>.05/.25</td>
<td>.06/.11</td>
<td>.01/.69</td>
<td>.81/.81</td>
<td>.3/.57</td>
</tr>
<tr>
<td>Percentage of words that are included in lexical chains with</td>
<td>.59/.68</td>
<td>.17/.7</td>
<td>.25/.36</td>
<td>.07/.35</td>
<td>.05/.36</td>
<td>.25/.34</td>
<td>.23/.47</td>
</tr>
<tr>
<td>more than 5 concepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discourse Factors**

<table>
<thead>
<tr>
<th>Factor</th>
<th>C1(%)</th>
<th>C2(%)</th>
<th>C3(%)</th>
<th>C4(%)</th>
<th>C5(%)</th>
<th>C6(%)</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average block score</td>
<td>.66/.86</td>
<td>.26/.66</td>
<td>.02/.29</td>
<td>.17/.25</td>
<td>.12/.78</td>
<td>.79/.84</td>
<td>.34/.61</td>
</tr>
<tr>
<td>Overall document score</td>
<td>.39/.66</td>
<td>.4/.73</td>
<td>.24/.55</td>
<td>.06/.47</td>
<td>.22/.33</td>
<td>.08/.25</td>
<td>.23/.5</td>
</tr>
<tr>
<td>Average block-document cohesion</td>
<td>.68/.93</td>
<td>.29/.72</td>
<td>.02/.36</td>
<td>.22/.25</td>
<td>.03/.75</td>
<td>.81/.83</td>
<td>.34/.64</td>
</tr>
<tr>
<td>Average sentence-block cohesion</td>
<td>.75/.84</td>
<td>.18/.58</td>
<td>.03/.35</td>
<td>.32/.38</td>
<td>.04/.79</td>
<td>.76/.8</td>
<td>.34/.62</td>
</tr>
<tr>
<td>Average inter-block cohesion</td>
<td>.51/.75</td>
<td>.22/.58</td>
<td>.01/.26</td>
<td>.1/.15</td>
<td>.05/.71</td>
<td>.81/.83</td>
<td>.28/.55</td>
</tr>
<tr>
<td>Average intra-block cohesion</td>
<td>.56/.75</td>
<td>.24/.62</td>
<td>.11/.39</td>
<td>.16/.38</td>
<td>.11/.74</td>
<td>.72/.79</td>
<td>.32/.61</td>
</tr>
</tbody>
</table>

### Appendix D – Input Examples


Ce soir-là, la famille de Matilda dînait comme d’habitude devant la télévision, quand ils entendirent une voix forte venant du salon dire : « salut, salut, salut ». La mère devint toute blanche. Elle dit à son mari « il y a quelqu’un dans la maison ». Ils arrêtèrent tous de manger. Ils étaient tous sur le qui-vive. La voix reprit « salut, salut, salut ». Le frère se mit à crier « ça recommence ! ». Matilda se leva et alla éteindre la télévision.

<< Verbalization breakpoint 1 >>

La mère, paniquée, dit à son mari : « Henri, des voleurs, ils sont dans le salon, tu devrais y aller ». Le père, raide sur sa chaise ne bougea pas. Il n’avait pas envie de jouer au héros.

Sa femme lui dit : « Alors, tu te décides ? Ils doivent être en train de faucher l’argenterie ! ».

<< Verbalization breakpoint 2 >>

Monsieur Verdebois s’essuya nerveusement les lèvres avec sa serviette et proposa d’aller voir tous ensemble. La mère attrapa un tisonnier au coin de la
cheminée. Le père s’arma d’une canne de golf posée dans un coin. Le frère attrapa un tabouret. Matilda prit le couteau avec lequel elle mangeait. Puis ils se dirigèrent tous les quatre vers la porte du salon en marchant sur la pointe des pieds.

À ce moment-là, ils entendirent à nouveau la voix. Matilda fit alors irruption dans la pièce en brandissant son couteau et cria « haut les mains, vous êtes pris ! ». Les autres la suivirent en agitant leurs armes.

<< Verbalization breakpoint 3 >>

Puis, ils s’arrêtèrent pour regarder autour d’eux. Ils ne virent personne. Le père fut soulagé et dit « il n’y a pas de voleur ici ». Sa femme lui répondit d’une voix tremblante « mais Henri, je l’ai entendu, et toi aussi ». Matilda appuya la réponse de sa mère en ajoutant « je suis sûre de l’avoir entendu, il est ici quelque part ».

C’est alors que la voix s’éleva à nouveau. Ils sursautèrent tous, y compris Matilda qui jouait très bien la comédie. Ils inspectèrent la grande pièce. Ils ne trouvèrent toujours personne.

<< Verbalization breakpoint 4 >>

Matilda dit alors que c’était un fantôme : « Le salon est hanté, je croyais que vous le saviez. Je sais que c’est le fantôme, je l’ai déjà entendu ici ». Les parents, très pâles, sortirent du salon suivis par les enfants.

<< Verbalization breakpoint 5 >>

Plus tard, suivie de son frère, Matilda retourna dans la pièce. C’est alors qu’elle sortit du manteau de la cheminée le perroquet de leur copain Arthur. Ils éclatèrent alors de rire. Ils passèrent par la porte de derrière en emmenant l’animal avec eux. Matilda rendit son perroquet à Arthur et lui raconta la soirée. Il n’y eut plus jamais de fantôme chez les Verdebois.

<< Verbalization breakpoint 6 >>
Sample Chat – Log of Team 4 Chat Conversation

<Dialog team="4" file="Team4.xml">
    <!-- predefined topics by the tutor -->
    <Topics>
        <Topic>Blog</Topic>
        <Topic>Chat</Topic>
        <Topic>Forum</Topic>
        <Topic>Wiki</Topic>
    </Topics>
    <Body>
        <Turn nickname="Participant 1">
            <Utterance genid="1" time="03.05.23" ref="0">joins the room</Utterance>
        </Turn>
        ...
        <!-- students select a technology out of the 4 initially suggested -->
        <Turn nickname="Participant 1">
            <Utterance genid="18" time="03.24.37" ref="0">I will tell you why my company loves blogs - in fact we have a product called Blog2007</Utterance>
        </Turn>
        ...
        <!-- students present arguments and debate the pros and cons of each technology -->
        <Turn nickname="Participant 4">
            <Utterance genid="27" time="03.26.02" ref="0">I think that a chat system for a company is much more suitable</Utterance>
        </Turn>
        ...
        <Turn nickname="Participant 1">
            <Utterance genid="41" time="03.28.49" ref="0">the major problem of wiki is that too many people can change the content</Utterance>
        </Turn>
    </Body>
</Dialog>
and so, it can be confusing

well this can be observed by somebody and not all changes are permanent

I've seen people telling how they solved catchy problems on blogs, talking techniques of programming and so on

the wiki is very professional

they were providing solutions, not talking about their personal life

4. The spread of chat is huge because of it's low bandwidth requirement

You can debate and get good answers, which can be explained right away if misunderstood

...
<!-- students combine the technologies in order to integrated environment -->

<Turn nickname="Participant 4">
  <Utterance genid="352" time="04.50.39" ref="0">using the wiki will furthermore take away from the work of the support team</Utterance>
</Turn>

<Turn nickname="Participant 3">
  <Utterance genid="353" time="04.50.42" ref="0">the forum is where this knowledge is discussed</Utterance>
</Turn>

…

<!-- end of conversation -->

<Turn nickname="Participant 1">
  <Utterance genid="395" time="04.57.43" ref="394">bye</Utterance>
</Turn>

<Turn nickname="Participant 1">
  <Utterance genid="396" time="04.57.54" ref="0">leaves the room</Utterance>
</Turn>

</Body>
</Dialog>

Sample Verbalization

The main reading strategies that were manually coded according to annotation methodology (see 8.1.3 Reading Strategies Identification Heuristics) were: Paraphrasing, Control, Bridging, Causality and Knowledge Inference. In addition, two other strategies were initially coded, but later on disregarded as they were insignificant as occurrences in terms of the overall self-explanations corpus: Generalization and Prediction (see Table 41).
**Table 41** Self-explanations example, manually coded in correspondence with the annotation methodology used by Nardy et al. (in press)

<table>
<thead>
<tr>
<th>No.</th>
<th>Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>Ben J’ai compris que c’est une famille la famille dans laquelle il ???suis/ qui dinent devant la télé et qui... tout d’un coup i z entendent une voix qui leur dit salut et du coup i z ont peur donc parce que la mère de Matilda ???donc c’est qué/ je pense qu’i z ont peur... alors i z arrêtent de manger puis le frère commence à comprendre quelque chose en disant ça recommence alors je pense je pense que c’est une famille peut-être assez riche parce qu’il y a de l’argenterie et qui pensent que ceux qui doit être riche ou qu’y a beaucoup de voleurs dans notre dans leur maison donc...</td>
</tr>
</tbody>
</table>
| V2  | donc la c’est... on sait déjà comment s’appelle la famille et puis ils racontent que là vu que le père veut pas y aller tout seul il est accompagné de toute sa famille pour... aller... voir s’y a un voleur et y a la le la parole /ça le bruit aussi ???/qui recommence *et du coup elle, la petite fille qui s’appelle Matilda commence à avoir peur (1) donc elle (1) lui (2) dit haut les mains vous êtes pris(2)*
| V3  | *Donc là... eee i i disent qu’y a pas de voleur et que par contre Matilda et sa maman sont sûres de l’avoir entendu et elle dit que y commencent à sursauter et que Matilda jouait très bien la comédie donc soit on peut penser que elle fait ça pour être d’accord avec ses parents ou soit que c’est elle qui a fait une blague ou soit que c’est peut-être une pièce... de théâtre ou quelque chose comme ça mais je pense surtout que c’est elle qui leur a fait une blague parce que... parce que... comme* |
| V4  | Donc là je confirme que c’est une blague par ce que (rire)... elle c’est elle qui éveille les soupçons donc... voilà |
| V5  | Donc là on sait que c’est une blague donc elle avait pris le perroquet de son copain pour jouer un tour à ses parents (E : tu vois autre chose à ajouter ?) Ben que... Par rapport à... |
Alias-i, LingPipe. 4.1.0 edn. (2008)


References


References


Glahn, C., Specht, M., Koper, R.: Reflecting on web-readings with tag clouds. Paper presented at the Computer-based Knowledge & Skill Assessment and Feedback in Learning Settings (CAF), Special Track at the 11th Int. Conf. on Interactive Computer aided Learning, ICL, Villach, Austria (2008)


Hansen, C., Emin, V., Wasson, B., Mor, Y., Rodríguez-Triana, M.J., Dascalu, M.,
Ferguson, R., Perin, J.-P.: Towards An Integrated Model of Teacher Inquiry into
Student Learning, Learning Design, and Learning Analytics. In: Hernández-Leo, D.,
Springer, Heidelberg (2013)
Harrer, A., Hever, R., Ziebarth, S.: Empowering researchers to detect interaction patterns in
e-collaboration. In: Luckin, R., Koedinger, K.R., Greer, J.E. (eds.) 13th Int. Conf. on
Intelligence and Applications, Los Angeles (2007)
Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer,
New York (2009)
Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi, M.: Classroom success of an
intelligent tutoring system for lexical practice and reading comprehension. In: 9th Int.
Conf. on Spoken Language Processing, p. 4. ISCA, Pittsburgh (2006)
Heinrich, G.: Parameter estimation for text analysis. vsonix GmbH + University of Leipzig,
Leipzig (2008)
Hirst, G., St-Onge, D.: Lexical Chains as representation of context for the detection and
Hmelo-Silver, C.E., Chernobilsky, E., Masto, O.: Analyzing collaborative learning:
Multiple approaches to understanding processes and outcomes. In: 7th Int. Conf. of the
Learning Sciences (ICLS 2006), Bloomington, IN, USA, pp. 1061–1062 (2006)
Hmelo-Silver, C.E., Chernobilsky, E., Nagarajan, A.: Two sides of the coin: Multiple
perspectives on collaborative knowledge construction in online problem-based learning.
Paper presented at the European Association for Research on Learning and Instruction,
Nicosia, Cyprus, August 23-27 (2005)
Hobbs, J.R.: On the Coherence and Structure of Discourse. Stanford University, Center for
the Study of Language and Information (1985)


Hsu, C.W., Chang, C.-C., Lin, C.-J.: A practical guide to support vector classification. National Taiwan University, Taipei (2010)


References


References


References


Miller, G.A.: The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review 63(2), 81–97 (1956)


Nguyen, Q.H., Hong, S.-H.: Comparison of centrality-based planarisation for 2.5D graph drawing. NICTA technical report, Sidney (2006)


Renaissance Learning, Matching Books to Students: How to Use Readability Formulas and Continuous Monitoring to Ensure Reading Success. Renaissance Learning, Inc., Wisconsin Rapids, WI (2011)


Tapiero, I.: Situation models and levels of coherence. Erlbaum, Mahwah (2007)


References


The first impression that the book may give to readers is its multiple threads of ideas and concepts. A main deal is about *textual cohesion*, which semantically glues words and sentences to each other. This book explores many ways of using cohesion to assess both learners’ input (course texts to read) and output (written productions, like summaries, forum discussion, or self-explanations). In a parallel way, as if Mihai Dascalu had used *ReaderBench* to get textual information about his book, we should emphasize the *high cohesiveness* of the book, since almost each theoretical sub-section has an echo in the empirical part.

*Echoes* are another important dimension of the book. They are related to cohesion, because this is induced by repetition. Moreover, echoes are an important ingredient of the polyphonic model that Mihai Dascalu used in the tools he developed. If we discuss about echoes, we should also consider the voices they originate from. Among these voices and their echoes appears the polyphonic framework, which characterizes complex coherent texts. *Complexity* is another fundamental concern in the book and might be considered in opposition with cohesion at a first sight. However, cohesion and complexity are not contradicting each other: a complex text may be or not cohesive and *vice versa*.

Since cohesion has to deal with semantic *connection* between words, we can argue that this book proves that one of the main abilities a PhD student has to develop is to take profit from the lab(s) environment(s), the work connections of his supervisor(s), and the multitude of projects they are in. Mihai Dascalu, as our former PhD student, has taken full advantage from these connections during the three-year long doctoral work, distributed between Bucharest (2 thirds of his time) and Grenoble (1 third of his time).

Mihai Dascalu's work has perfectly captured all the features of the concepts he has studied along his PhD research: multi-vocal, polyphonic, complex, and coherent. He succeeded to integrate multiple concepts from different domains in a polyphonic style: education theories, psychology, linguistics, philology, music, computer science, and artificial intelligence. This book will surely contribute to a better understanding and a spreading of how texts to be learnt can be data-mined for providing learners with immediate support and feedback.

Philippe Dessus  
University Grenoble Alpes, France  
Stefan Trausan-Matu  
University “Politehnica” of Bucharest, Romania