HAL
open science

ReaderBench, o platformă integrată pentru analiza complexității textuale și a strategiilor de lectură

Mihai Dascălu, Ștefan Trăușan-Matu, Philippe Dessus, Maryse Bianco, Aurélie Nardy

To cite this version:

Mihai Dascălu, Ștefan Trăușan-Matu, Philippe Dessus, Maryse Bianco, Aurélie Nardy. ReaderBench, o platformă integrată pentru analiza complexității textuale și a strategiilor de lectură. Proc. 10-a Conf. Nat. de Interactiune Om-Calculator (RoCHI 2013), T. Stefanut; C. Rusu, 2013, Cluj, Romania. pp.39-46. hal-01412573

HAL Id: hal-01412573

https://hal.science/hal-01412573

Submitted on 8 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ReaderBench, o platformă integrată pentru analiza complexității textuale și a strategiilor de lectură

Mihai Dascălu, Ștefan Trăușan-Matu
Universitatea Politehnica din București
313 Splaiul Independenței, 060042 București, România
mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro

Philippe Dessus, Maryse Bianco, Aurélie Nardy
LSE, Université Pierre-Mendès France
F-38040 Grenoble CEDEX 9, Franța
philippe.dessus, maryse.bianco, aurelie.nardy@upmf-grenoble.fr

REZUMAT

ReaderBench este o platformă integrată care poate fi utilizată pentru diverse scopuri didactice. Ea permite evaluarea unei game largi de producții textuale ale cursanților și utilizarea lor de către profesori, asigurând totodată suport multilingv și flexibilitate. *ReaderBench* permite evaluarea a trei caracteristici principale ale textelor: coeziunea între diverse fragmente de text, identificarea strategiilor de lectură și evaluarea complexității textuale. Fiecare dintre aceste dimensiuni ale analizei au făcut obiectul unor validări empirice. *ReaderBench* acoperă un ciclu complet de învățare, de la evaluarea inițială a complexității materialelor de lectură, alocarea textelor cursanților, surprinderea meta-cognițiilor reflectate în verbalizările textuale și evaluarea înțelegerii, încurajând, prin urmare, procesul de autoreglare a cursantului.

Cuvinte cheie

coeziunea textelor, strategii de lectură, complexitatea textuală, analiza semantică latentă (LSA – Latent Semantic Analysis), alocarea latentă Dirichlet (LDA – Latent Dirichlet Allocation), mașini cu vector suport (SVM – Support Vector Machine).

Clasificare ACM

I.2.7 Natural Language Processing.

INTRODUCERE

În multe situații de instruire, lectura de materiale textuale și reflectarea la cele citite („auto-explicarea”) urmată uneori de transcrierea gândurilor sunt activitățile de bază care determină învățarea (din punctul de vedere al cursantului) și indicatori ai procesului de învățare (din punctul de vedere al profesorului). Lectura este o activitate cognitivă ale cărei urmări orale sau scrise sunt de obicei analizate de către profesori pentru a deduce înțelegerea cursanților. O posibilă alternativă în vederea evaluării nivelului de înțelegere al elevilor presupune evaluarea strategiilor de lectură utilizate. Prin urmare, citirea și scrierea sunt activitățile de bază pe care fiecare profesor trebuie să le evalueze zi de zi: materialele de lectură trebuie să fie întocmite sau adaptate pentru a fi în conformitate cu nivelul actual al cursanților, iar strategiile de lectură trebuie să fie analizate pentru a deduce nivelul cursanților de procesare și de înțelegere a textului.

Profesorilor le este dificil să ofere sprijin cursanților pe parcursul activităților lor de citire și scriere pe o scară mai largă, prin urmare, ei trebuie să se ocupe de un număr mic de elevi. Mai mult decât atât, evaluarea materialelor textuale și a verbalizărilor este o activitate solicitantă din punct de vedere cognitiv, subiectivă prin însăși dificultatea definirii unor criterii de evaluare obiective și suficient de generale. În acest context, am proiectat și implementat *ReaderBench* [3, 6], un mediu flexibil asistat de calculator care susține activitățile de lectură și scriere ale cursanților, precum și cele de evaluare ale cadrelor didactice, în cadrul a multiple scenarii educaționale.

Următoarea secțiune detaliază unii dintre principalii predictori de înțelegere a lecturii care au condus la introducerea sistemului *ReaderBench*. A treia secțiune este centrată pe analiza coeziunii textuale, considerată un element esențial în analiza discursului. Apoi ne-am îndreptat atenția spre strategiile de lectură și evaluarea complexității textuale. Fiecare dintre ultimele trei secțiuni este însoțită de o validare a *ReaderBench* raportat la evaluări umane.

PREDICTORI DE BAZĂ AI ÎNȚELEGERII LECTURII

Cititorii experți sunt cititori strategici. Ei își monitorizează lectura, fiind mereu capabili să-și cunoască în fiecare moment nivelul lor de înțelegere. Atunci când aceștia se confruntă cu o dificultate, pot apela la procedurile de autoreglare, numite strategii de lectură [24]. Acestea au fost studiate extensiv cu ajutorul cititorilor adolescenți și adulți care utilizează procedura lecturii cu voce tare, care determină cititorul să-și auto-explice în momentele de oprire specifice în timpul lecturii, prin urmare, oferind o explicație în ceea ce privește înțelegerea.

Patru tipuri de strategii de lectură sunt folosite în principal de către cititorii experți [20]. *Parafrazarea* permite cititorului să exprime ceea ce a înțeles din conținutul explicit al textului și poate fi considerat primul pas esențial în procesul de construire a coerenței. *Inferențele bazate pe text*, de exemplu strategiile de cauzalitate și de corelare, construiesc relații explicite între două sau mai multe fragmente de texte. Pe de altă parte, *inferențele bazate pe cunoaștere* construiesc relații între informațiile din text și cunoștințe proprii ale cititorului și sunt esențiale pentru construirea unui model situațional [34]. *Strategiile de control* se referă la procesul de monitorizare propriu-zisă, atunci când cititorul exprimă în mod explicit ceea ce a înțeles sau nu. Diversitatea și

bogăția strategiilor pe care un cititor le stăpânește depind de mai mulți factori, fie personali (competență, nivelul de cunoștințe, motivație), sau externi (complexitatea textuală a materialelor utilizate).

În plus, profesorii au nevoie de măsuri de complexitate textuală valide și de încredere pentru selectarea textelor în cadrul procesului educațional de zi cu zi. Două abordări concură în evaluarea automată a complexității textului: 1/ folosirea de metrici simple, statistice care se bazează în mare parte pe dificultatea cuvântului (din statistici efectuate în prealabil pe corpusuri de texte) și lungimea propoziției, 2/ folosirea unei combinații de mai mulți factori, variind de la indicatori lexicali, precum frecvența cuvintelor, până la niveluri sintactice și semantice (de exemplu, coeziunea textuală) [28].

Ca perspectivă, coeziunea textului, văzută ca similaritate între diferite părți ale textelor este un factor determinant major de coerență a textului și s-a dovedit a fi un predictor important de înțelegere a lecturii [31]. Înțelegerea coeziunii (de exemplu, referențială, de cauzalitate sau temporală) este esențială pentru procesul de construire a coerenței unui text la nivel local, care, la rândul său, permite reorganizarea conținutul textual la nivel de macrostructură, respectiv model situațional. Textele cu coeziune ridicată sunt mai benefice pentru cititorii puțin cunoscători decât pentru cititorii cu cunoștințe aprofundate [23]. Prin urmare, coeziunea textuală este o caracteristică de complexitate textuală care reliefează caracteristici semantice desprinse din textul citit și care poate interfera cu strategiile de lectură prin deducțiile realizate de către cititor.

McNamara și colegii săi au conceput două sisteme: în timp ce *CohMetrix* [11, 23] abordează aspecte ale complexității textuale, *iStart* [21, 25] este axat pe strategii de lectură. *CohMetrix* oferă o gamă largă de factori privind caracteristicile textuale, organizați pe cinci niveluri principale: cuvânt (de exemplu, parte de vorbire și frecvență de apariție aferentă), sintaxă (de exemplu, procentul de substantive), baza textului care reflectă ideile explicite (de exemplu, co-referințe și diversitate lexicală), modelul situațional (de exemplu, coeziunea și indicii temporali), și gen și structură retorică (de exemplu, genul textului) [10].

iStart este primul sistem implementat care învață și evaluează auto-explicațiile cursanților în conformitate cu materialul citit, cu ajutorul a diferite module bazate pe Self-Explanation Reading Training (SERT) [20]. Sistemul este compus din două module principale: un modul este centrat pe prezentarea tehnicilor SERT prin utilizarea unui student virtual, în timp ce un alt modul cere cursanților să citească textele, înregistrează verbalizările acestora, le evaluează și oferă feedback personalizat.

ReaderBench integrează funcționalitățile *CohMetrix* și *iStart*, deoarece oferă cadrelor didactice și elevilor informații despre activitățile lor de citire/scriere: evaluarea inițială a complexității textuale, atribuirea de texte cursanților, surprinderea meta-cognițiilor reflectate în verbalizările textuale și evaluarea strategiilor de lectură. Principalele diferențe între *ReaderBench* și sistemele anterioare constau în următoarele: 1/ utilizarea unui model generalizat al discursului bazat pe coeziune, care poate fi extins cu ușurință, de la analiza eseurilor cursanților sau povești, până la analiza discuțiilor și a forumurilor, cu accent pe evaluarea colaborării [6, 33]; 2/ factori diferiți, măsurători și utilizarea clasificatorilor bazați pe metode kernel – SVM (Support Vector Machine) pentru creșterea validității evaluării complexității textuale [5]; 3/ suport multi-lingual și integrarea de instrumente de prelucrare a limbajului natural specifice atât pentru limba franceză, cât și pentru engleză; 4/ un scop educativ diferit, pentru că validarea sistemului *ReaderBench* a fost efectuată pe elevi (din clasa a 3-a până întra 5-a), în timp ce *iStart* vizează în principal elevii de liceu și studenții.

Mai mult, proiectarea *ReaderBench* a luat în considerare două dimensiuni. Pe de o parte, *flexibilitatea* mediului este evidențiată prin următoarele caracteristici: comparația între nivelurile de complexitate a mai multor texte, și ușurința de editare a materialelor de lectură din cadrul *ReaderBench*, cu posibilitatea de a adăuga, de asemenea, momente dinamice pentru a permite cursanților să-și exprime verbalizările sau rezumatele. Profesorii pot gestiona astfel materiale textuale în vederea ajungerii la caracteristicile dorite. De asemenea, cursanții își pot face foarte repede o idee asupra modului în care aceștia își pot auto-regla lectura prin evaluarea proprie a strategiilor utilizate. Pe de altă parte, *extensibilitatea* sistemului se reflectă în ușurința antrenării și folosirii de spații adiționale pentru analiza semantică latentă (LSA – Latent Semantic Analysis) [13], respectiv modele de topice aferente alocării latente Dirichlet (LDA – Latent Dirichlet Allocation) [1], sau în posibilitatea de a adăuga noi factori în vederea evaluării complexității textuale.

ANALIZA DISCURSULUI BAZAT PE COEZIUNE

Coeziunea textului, privită din prisma relațiilor lexicale, gramaticale și semantice care definesc structura textului, este compusă în modelul nostru implementat din: 1/ inversul distanței între elementele textual analizate; 2/ proximitatea lexicală identificată prin lemele identice ale cuvintelor și distanțele semantice [2, 15, 35] din cadrul ontologiilor; 3/ similaritatea semantică măsurată prin LSA [13] și LDA [1]. În plus, tehnici specifice prelucrării limbajului natural sunt aplicate pentru a reduce zgomotul și pentru a îmbunătăți precizia sistemului: selecția cuvintelor existente în dicționar, stemming, lematizare, recunoașterea entităților numite, adnotarea cu părți de vorbire, parsarea și rezoluția co-referențială [18].

Pentru a oferi o platformă de analiză multi-lingvistică pentru limbile engleză și franceză, *ReaderBench* integrează atât WordNet [26], cât și o versiune serializată a WOLF (WordNet Libre du Français, http://alpage.inria.fr/~Sagot/wolf.html) [30]. Datorită limitărilor intrinsece ale WOLF, în care conceptele sunt traduse din limba engleză în timp ce glosele lor corespunzătoare sunt doar parțial traduse, făcând un amestec de definiții din franceză și engleză, doar trei distanțe semantice utilizate frecvent au rămas aplicabile ambelor ontologii: lungimea căii, Wu–Palmer [35] și Leacock–Chodorow's (lungimea căii normalizată) [15].

Ulterior, modelele semantice au fost antrenate folosind trei corpusuri specifice: "TextEnfants" [7] (aproximativ 4,2

milioane de cuvinte), "Le Monde" (ziar francez, aproximativ 24 milioane de cuvinte) pentru limba franceză, și corpusul Touchstone Applied Science Associates (TASA) (aproximativ 13 milioane de cuvinte) pentru limba engleză. Mai mult decât atât, au fost aplicate îmbunătățiri modelelor inițiale: reducerea formelor flexionare la lemele lor, adnotarea fiecărui cuvânt cu partea corespunzătoare a vorbirii, cu ajutorul NLP, normalizarea apariției cuvintelor prin folosirea măsurii „frecvența conceptelor-inversul frecvenței documentelor" (Tf-Idf term frequency-inverse document frequency) [18] și calcul distribuit pentru creșterea vitezei de procesare [4, 19].

Modelele LSA și LDA extrag relații semantice din co-aparițiile conceptelor și se bazează pe ipoteza sac-de-cuvinte [18]. Experimentele noastre au dovedit că modelele LSA și LDA pot fi folosite pentru a se completa reciproc, în sensul că relațiile semantice, coezive, care stau la bază sunt mai susceptibile de a fi identificate, în cazul în care ambele abordări sunt combinate după normalizare. Prin urmare, vectorii LSA sunt generați după proiecția vectorilor obținuți în urma descompuneri în valori singulare (SVD – Singular Value Decomposition) a matricei Tf-Idf inițială. Vectorii rezultați pot fi utilizați pentru a determina proximitatea cuvintelor prin aplicarea similarității de tip cosinus [14]. Din alt punct de vedere, modelele de topice LDA oferă un mecanism de inferență a structurii topicelor (clase de concepte) printr-un proces probabilistic generativ [1]. În acest context, similaritatea dintre concepte poate fi văzută ca opusul divergenței Jensen-Shannon [18] între distribuțiile a posteriori raportată la topicele considerate.

În general, în scopul de a înțelege mai bine coeziunea între fragmentele textuale, am combinat tehnici specifice de recuperare a informațiilor [17], reflectate mai ales în repetițiile de cuvinte și numărul normalizat de apariții, cu distanțe semantice extrase din ontologii sau din modelele semantice bazate pe LSA sau LDA.

Pentru a avea o mai bună reprezentare a discursului în ceea ce privește legăturile coezive care stau la baza sa, am introdus un grafic de coeziune, care poate fi văzută ca o generalizare a graficului propus anterior de intervenții [32]. Astfel, un graf mixt, multi-nivel este generat alcătuit din trei tipuri de noduri: un nod central, *documentul* care reprezintă întregul material de citit, noduri pentru fiecare *bloc de text* sau paragrafe din materialul inițial și pentru *propoziții*, principalele unități de analiză. Drept restricții de construcție, legăturile ierarhice iau în considerare funcțiile de incluziune (propozițiile în cadrului într-un bloc de text, blocuri în cadrul documentului per ansamblu), precum și două tipuri de legături introduse între elementele de analiză din cadrul aceluiași nivel. Legăturile obligatorii sunt stabilite între paragrafele sau propoziții adiacente și sunt folosite pentru a modelare coerent fluxul de informații pe tot parcursul discursului, făcând posibilă identificarea pauzelor de coeziune. Legături suplimentare relevante sunt adăugate în cadrul grafului de coeziune pentru a evidenția relațiile fine dintre elementele de analiză îndepărtate. În experimentele noastre, utilizarea sumei dintre valoarea medie și deviația standard a tuturor valorilor de coeziune drept prag de coeziune a oferit legături suplimentare semnificative în structura discursului analizat.

În schimb, de vreme ce coeziunea poate fi considerată ca o sumă a legăturilor care țin un text împreună și îi dau sens, simpla utilizare a cuvintelor legate semantic într-un text nu corelează direct cu complexitatea acestuia. Cu alte cuvinte, în timp ce coeziunea în sine nu este suficientă pentru a distinge un text în ceea ce privește complexitatea aferentă, lipsa coeziunii poate crește complexitatea textuală, întrucât înțelegerea și reprezentarea corectă a textului devin mai dificil de realizat. Pentru a evidenția mai bine acest punct de vedere, au fost definite două măsuri pentru complexitatea textuală, detaliate ulterior: coeziunea din interiorul blocului ca valoarea medie a tuturor legăturilor din cadrul unui bloc (legături adiacente și relevante între propoziții) și coeziunea dintre blocurile de text, care evidențiază relațiile semantice la nivel de document global.

Pentru validarea funcției de coeziune, am folosit 10 de povești în limba franceză, pentru care studenții din anul al doilea de la Științele Educației (vorbitori nativi de limba franceza) au fost rugați să evalueze înrudirea semantică între paragrafe adiacente pe o scală Likert de [1..5]. Fiecare pereche de paragrafe a fost evaluată de mai mult de 10 evaluatori umani pentru o limitare a dezacordului între evaluatori. Având în vedere subiectivitatea sarcinii și diferențele de percepție personală ale coeziunii, valorile medii ale corelațiilor din cadrul aceleiași clase pentru fiecare poveste au fost: ICC (intra-class correlation) raportată la valoarea medie = .493 și ICC raportată la măsuri singulare = .167. În final, 540 de valori individuale de coeziune au fost agregate și apoi utilizate pentru a determina corelația dintre diferite măsuri semantice, și standardul de aur. Pentru cele două corpusuri folosite (Le Monde și TextEnfants), corelațiile au fost [3]: combinat-Le Monde (r = .54), LDA-Le Monde (r = .42), LSA-Le Monde (r = .28), LSA-TextEnfants (r = .19), combinat-TextEnfants (r = .06), Wu-Palmer (r = -.06), Path Similarity (r = -.13), LDA-TextEnfants (r = -.13) și Leacock-Chodorow (r = -.40).

Rezultatele anterioare arată că metoda propusă de a combina mai multe măsuri de similaritate semantică surclasează toate măsurătorile individuale, că un corpus mai mare conduce la rezultate mai bune și că distanța Wu-Palmer, pe lângă scalarea corespunzătoare în intervalul [0, 1] (relevantă la integrarea cu LSA și LDA), se comportă cel mai bine în contrast cu alte distanțe semantice bazate pe ontologii. În plus, creșterea semnificativă a corelației între măsura agregată a LSA, LDA și Wu-Palmer, în comparație cu scorurile individuale, demonstrează beneficiile combinării mai multor abordări complementare în ceea ce privește reducerea erorilor care pot fi induse prin utilizarea unei singure metode.

STRATEGII DE LECTURĂ

Pornind de la cele patru tipuri de strategii de lectură prezentate în secțiunea a 2-a, scopul nostru a fost de a integra metode de extracție automată concepute pentru a sprijini profesorii în identificarea diverselor strategii de lectură, asigurând totodată o aliniere cât mai bună cu

categoriile SERT [20]. În acest context am testat diverse metode de identificare a strategiilor de lectură (cauzalitate, control, parafrazare, corelare, și interferența de cunoștințe); în cadrul articolului ne vom concentra exclusiv asupra prezentării alternativelor care au oferit cele mai bune corelări globale om-mașină.

Raportat la complexitatea procesului de identificare, cele mai simple strategii sunt *cauzalitatea* (de exemplu, fr., "parce que", "pour") și *controlul* (de exemplu, fr., "Je me souviens", "Je crois"), pentru care s-au folosit sintagme predefinite de identificare („cue phrases"). În plus, odată ce cauzalitatea presupune inferențe bazate pe text, s-a renunțat la apariția de cuvinte cheie la începutul unei verbalizări întrucât aceasta poate fi considerată un eveniment de inițiere a unui discurs (de exemplu, "Donc"), mai degrabă decât unul de creare a unei legături inferențiale. Ulterior, *parafrazele* care au fost considerate repetări ale aceloraşi propuneri semantice de către evaluatori, au fost identificate automat pe baza lemelor cuvintelor și pe baza relațiilor de sinonimie din ontologiile lexicalizate.

Strategiile cel mai greu de identificat sunt *inferența de cunoștințe* și *corelarea*, pentru care similarități semantice trebuie să fie calculate. Un concept dedus este un cuvânt non-parafrazat pentru care s-au calculat următoarele trei distanțe semantice: distanța de la cuvântul W_1 de la verbalizare la cel mai apropiat cuvânt W_2 din textul inițial (exprimat în termeni de distanțe semantice în ontologii, LSA și LDA) și distanțele de la W_1 și W_2 la fragmentele textuale între auto-explicații consecutive. Distanțele din urmă au trebuit să fie luate în considerare pentru a pondera cât mai bine importanța fiecărui concept, în raport cu text integral.

Întrucât corelarea constă în crearea de legături între diferite segmente de text din textul inițial, coeziunea a fost măsurată între verbalizare și fiecare propoziție din materiale de citit de referință. În cazul în care mai mult de două măsuri de similaritate au fost peste valoarea medie a tuturor similarităților semantice anterioare și au depășit un prag minim, corelarea a fost estimată ca numărul de legături de coeziune între zone contigue de propoziții coerente. Aceasta a fost o adaptare în ceea ce privește procesul de adnotarea manuală care a luat în considerare două sau mai multe propoziții adiacente, fiecare coezive cu verbalizarea, membre a unei singure entități unite de corelarea cu textul inițial.

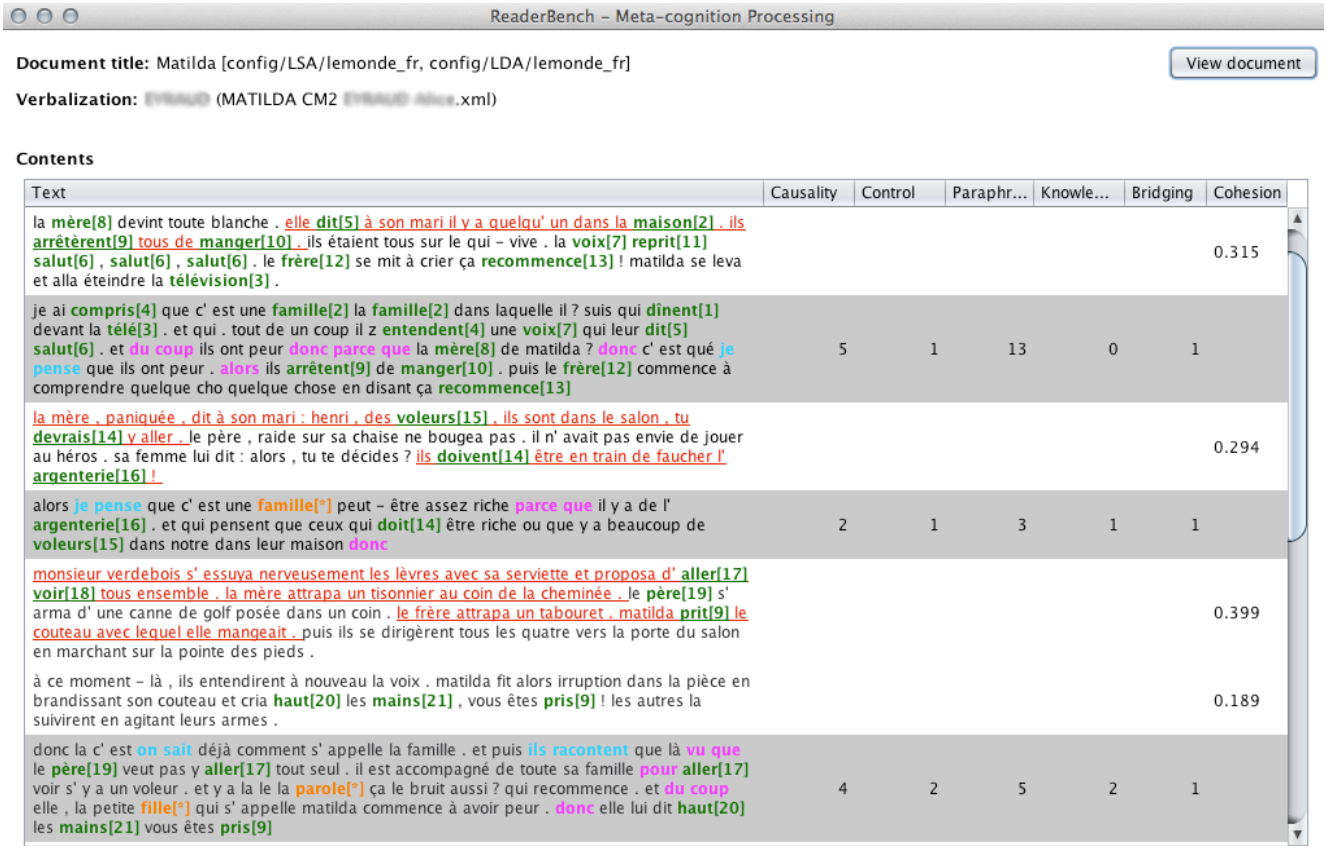


Figura 1. Strategiile de lectură identificate automat în cadrul verbalizării elevilor (gri) în cadrul ReaderBench [3]

Figura 1 prezintă măsurile de coeziune din paragrafele anterioare din poveste în ultima coloană și strategiile de lectură identificate pentru fiecare verbalizare marcate în zonele gri, codificate după cum urmează: control, cauzalitate, parafrazare [indexul raportat la cuvântul menționat din textul inițial], concept dedus [*] și corelare peste propozițiile coezive interconectate din materialul de citit.

În vederea validării metodei, am desfășurat un experiment cu elevi cu vârste cuprinse între 9 și 11 ani, care au trebuit să citească cu voce tare o poveste de 450 de cuvinte, să se oprească la șase markeri predefiniți și să explice ceea ce au înțeles până în acel moment. Explicațiile lor au fost mai întâi înregistrate și transcrise, apoi adnotate de către doi experți umani (care dețin titlul de doctor în lingvistică, respectiv doctor în psihologie), și clasificate în funcție de

sistemul lui McNamara [20]. Dezacordurile au fost soluționate prin discuții după evaluarea fiecărei auto-explicații individuale. În plus, o curățare automată a trebuit să fie efectuată pentru a putea procesa automat verbalizările transcrise fonetic. Verbalizările a 12 elevi au fost transcrise și evaluate manual drept validare preliminară. Rezultatele pentru cele 72 de auto-explicații în ceea ce privește precizia, rechemarea și scorul F1 sunt după cum urmează: cauzalitatea ($P = .57$, $R = .98$, $F = .72$), control ($P = 1$, $R = .71$, $F = 0.83$), parafrazare ($P = .79$, $R = .92$, $F = .85$), deducerea de cunoștințe ($P = .34$, $R = .43$, $F = .38$) și corelarea ($P = .45$, $R = .58$, $F = .5$). Cum era de așteptat, parafrazele, controlul și cauzalitatea au fost mult mai ușor de identificat decât informațiile care se raportau la experiența elevilor [12].

Mai mult decât atât, am identificat mai multe cazuri particulare, în care ambele abordări (umane și automate) fac obiectul unui adevăr parțial, care este în final subiectiv raportat la evaluator. De exemplu, multe structuri cauzale apropiate una de cealaltă, dar nu adiacente, au fost codificate manual ca fiind una, în timp ce sistemul le-a luat în considerare pe fiecare dintre ele separat. Mai mult decât atât, conceptul "fille" nu apare în text și este direct legat de personajul principal al textului, prin urmare marcat drept concept dedus de către *ReaderBench*, în timp ce evaluatorul l-a considerat un sinonim. Cu toate acestea, scopul nostru a fost de a sprijini profesorii și rezultatele sunt încurajatoare (fapt corelat și cu măsurătorile anterioare de precizie și cu faptul că a existat o mulțime de zgomot în transcrieri), subliniind beneficiile unui proces regularizat și determinist de identificare.

COMPLEXITATEA TEXTUALĂ

Evaluarea complexității textuale poate fi considerată o sarcină dificilă din cauza percepțiilor diferite ale cititorului cauzate în primul rând de cunoștințele anterioare și experiența, capacitatea cognitivă, motivația, interesele sau familiaritatea limbii (pentru non-vorbitori nativi). Cu toate acestea, din punct de vedere al profesorului, sarcina de a identifica materiale accesibile curanților joacă un rol crucial în procesul de învățare întrucât textele inadecvate, fie prea simple sau prea dificile, pot determina ca elevii să-și piardă rapid interesul. Noi propunem astfel o analiză multi-dimensională a complexității textuale, acoperind o multitudine de factori, prezentați în tabelul 1 (descrierea extinsă a acestora se regăsește în [5]) agregate prin utilizarea SVM, care s-a dovedit a fi cea mai eficientă [27] pentru că variabilele nu sunt liniar separabile.

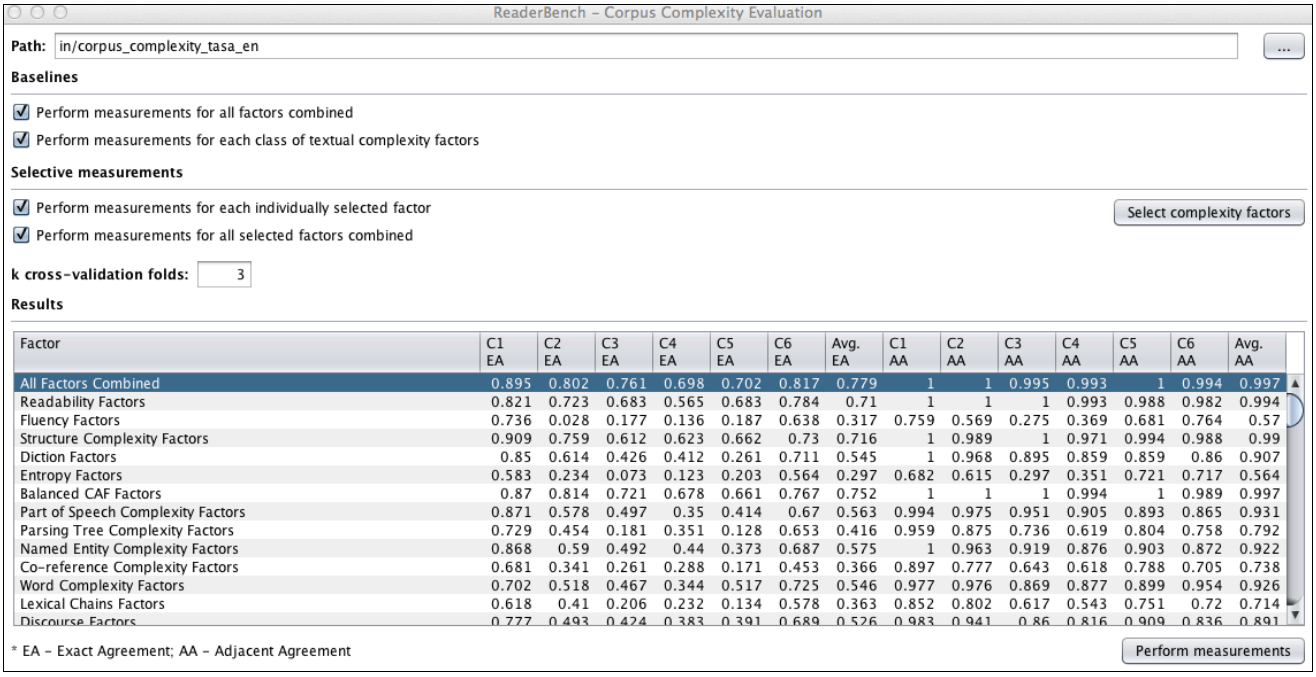
Tabelul 1. Dimensiunile analizei complexității textuale [3]

Adâncimea analizei	Factorii de evaluare	AE Mediu	AA Mediu
Suprafață	Formule de lizibilitate	.717	.995
	Factori de fluență	.314	.579
	Factori de complexitate a structurii textului	.728	.993
	Factori de dicție	.550	.901
	Entropie (la nivelul cuvintelor vs. caractere)	.313	.573
	Factori de complexitate ai cuvintelor	.556	.918
Morfologie și Sintaxă	Metrica CAF (Complexitate, Acuratețe, Fluență) balansată	.755	.996
	Factori de complexitate specifici părților de vorbire	.570	.929
	Complexitatea arborelui de parsare	.424	.806
Semantică	Coeziunea obținută prin lanțuri lexicale, LSA și LDA	.544	.894
	Factori de complexitate desprinși din statistici aplicate entităților numite	.590	.929
	Factori de complexitate co-referențială	.384	.730
	Factori bazați pe lanțuri lexicale identificate	.367	.704

Prin urmare, în afară de factorii de suprafață prezentați în [5], de mare interes este modul în care factorii morfologici și semantici se corelează cu măsuri de lizibilitate clasice. Prin urmare, pornind de la modelul de complexitate textuală care a integrat deja aceste măsuri, metrici de suprafață provenite din tehnicile automate de notare a eseurilor [29], factorii de morfologie și sintaxă [5], am introdus în cadrul modelului integrat de evaluare a complexității textuale noi dimensiuni axate pe semantică. În primul rând, coeziunea reflectată în puterea legăturilor din interiorul blocului de text și între blocuri influențează lizibilitatea, întrucât similaritățile semantice guvernează înțelegerea unui text. În al doilea rând, o varietate de metrici bazate pe lungimea și acoperirea lanțurilor lexicale [8] oferă o perspectivă în ceea ce privește varietatea lexiconului și coeziunea textului. În al treilea rând, caracteristicile bazate pe densitatea entităților numite s-au dovedit a influența lizibilitatea odată ce numărul de entități introduse într-un text este corelat cu memoria de lucru a cititorilor vizați ai textului. În cele din urmă, o altă dimensiune se concentrează pe capacitatea de a rezolva relațiile referențiale în mod corect [16] de vreme ce caracteristicile de inferență co-referențială influențează, de asemenea, aprecierea dificultății unui text. Dintr-o altă perspectivă, complexitatea cuvintelor utilizate a fost tratată ca o combinație dintre următorii factori: numărul de silabe, distanța dintre forma flexionară, lemă și stem, iar specificitatea este reflectată prin frecvența inversă de apariție în documentele din corpus, distanța în cadrul arborelui de hipernimie și numărul de sensuri posibile ale cuvintelor în cadrul ontologiei.

În vederea validării mecanismului de predicție a complexității unui text am optat pentru a extrage automat textele din corpusul TASA folosind scorul Degree of

Reading Power (DRP) şi clasificarea documentelor în şase clase de complexitate [22] de frecvenţă egală, necesare pentru realizarea clasificării pe baza SVM-urilor. Acest scenariu de validare care a presupus utilizarea a 1.020 de documente a avut un dublu scop: pe de o parte că modelul complet este adecvat şi fiabil şi, pe de altă parte, că metricile de complexitate nivel înalt (morfologic şi semantic) oferă o perspectiva relevantă, care poate fi utilizată pentru clasificarea automată a documentelor. Drept aspecte specifice de utilizare SVM-urilor, toţi factorii au fost scalaţi liniar şi a fost pusă în aplicare o metodă de optimizare de tip Grid Search a parametrilor *C* şi γ utilizaţi de către funcţia kernel de tip Gaussian. În cele din urmă, metoda k-fold cross validation [9] a fost aplicată pentru a extragere următoarele caracteristici de performanţă (a se vedea tabelul 1 şi Figura 2): precizia sau acordul exact (AE) şi acordul adiacent (AA) [27], ca procent la care SVM a fost aproape de a prezice clasificarea corectă. Astfel, prin intermediul interfeţei prezentate în Figura 2 utilizatorul are oportunitatea să realizeze următoarele măsurători: 1/ aplicarea modelului SVM complet cu toţi factorii de evaluare integraţi; 2/ identificarea perfomanţei fiecărei dimensiuni de analiză, pornind de la un set predefinit de metrici individuale de complexitatea textuală; 3/ aplicarea unui set specific de factori de complexitate textuală pentru simularea unui nou model de complexitate textuală.



Factor	C1 EA	C2 EA	C3 EA	C4 EA	C5 EA	C6 EA	Avg. EA	C1 AA	C2 AA	C3 AA	C4 AA	C5 AA	C6 AA	Avg. AA
All Factors Combined	0.895	0.802	0.761	0.698	0.702	0.817	0.779	1	1	0.995	0.993	1	0.994	0.997
Readability Factors	0.821	0.723	0.683	0.565	0.683	0.784	0.71	1	1	1	0.993	0.988	0.982	0.994
Fluency Factors	0.736	0.028	0.177	0.136	0.187	0.638	0.317	0.759	0.569	0.275	0.369	0.681	0.764	0.57
Structure Complexity Factors	0.909	0.759	0.612	0.623	0.662	0.73	0.716	1	0.989	1	0.971	0.994	0.988	0.99
Diction Factors	0.85	0.614	0.426	0.412	0.261	0.711	0.545	1	0.968	0.895	0.859	0.859	0.86	0.907
Entropy Factors	0.583	0.234	0.073	0.123	0.203	0.564	0.297	0.682	0.615	0.297	0.351	0.721	0.717	0.564
Balanced CAF Factors	0.87	0.814	0.721	0.678	0.661	0.767	0.752	1	1	1	0.994	1	0.989	0.997
Part of Speech Complexity Factors	0.871	0.578	0.497	0.35	0.414	0.67	0.563	0.994	0.975	0.951	0.905	0.893	0.865	0.931
Parsing Tree Complexity Factors	0.729	0.454	0.181	0.351	0.128	0.653	0.416	0.959	0.875	0.736	0.619	0.804	0.758	0.792
Named Entity Complexity Factors	0.868	0.59	0.492	0.44	0.373	0.687	0.575	1	0.963	0.919	0.876	0.903	0.872	0.922
Co-reference Complexity Factors	0.681	0.341	0.261	0.288	0.171	0.453	0.366	0.897	0.777	0.643	0.618	0.788	0.705	0.738
Word Complexity Factors	0.702	0.518	0.467	0.344	0.517	0.725	0.546	0.977	0.976	0.869	0.877	0.899	0.954	0.926
Lexical Chains Factors	0.618	0.41	0.206	0.232	0.134	0.578	0.363	0.852	0.802	0.617	0.543	0.751	0.72	0.714
Discourse Factors	0.777	0.493	0.424	0.383	0.391	0.689	0.526	0.983	0.941	0.86	0.816	0.909	0.836	0.891

Figura 2. Evaluarea complexităţii textuale

Mai mult decât atât, s-au efectuat două măsurători suplimentare. În primul rând, integrarea tuturor metricilor din toate clasele de complexitate a demonstrat că rezultatele SVM sunt compatibile cu scorurile DRP (AE = .763 şi AA = .997), şi că acestea oferă îmbunătăţiri semnificative, deoarece depăşesc precizia oricărei dimensiuni de factori de complexitate textuală considerată individual. A doua măsură (AE = .597 şi AA = .943) utilizează numai factori morfologici şi semantici pentru a evita o comparaţie circulară între factori de complexitate similară, de vreme ce scorul DRP se bazează pe factori exclusiv de suprafaţă. Acest rezultat arată o legătură între factorii de nivel scăzut (utilizaţi de asemenea în scorul DRP) şi factorii de analiză în profunzime, care pot fi de asemenea folosiţi pentru a prezice cu exactitate complexitatea unui material de lectură.

CONCLUZII ŞI DIRECŢII VIITOARE DE CERCETARE

ReaderBench este o platformă de integrare a noi modalităţi de evaluare a unei game largi de procese cognitive implicate în procesul de lectură, prin utilizarea unor tehnici avansate de prelucrare a limbajului natural. Acesta oferă o perspectivă semantică a analizei şi o structură de discurs prin combinarea de multiple distanţelor semantice. Flexibilitatea (suport multilingv) şi extensibilitatea sa (factori de complexitate uşor incorporabili) creşte aplicabilitatea sa în diverse contexte educaţionale (de exemplu, înţelegerea materialelor de lectură sau analiza fişelor de lectură). Totodată, prin utilizarea un corpus evaluat de către experţi pentru antrenarea modelului de complexitate textuală, *ReaderBench* va sprijini în mod eficient cursanţii în activităţile lor de învăţare. Mai mult decât atât, funcţionalitatea de transcriere speech-to-text, chiar dacă limitată de performanţele actuale ale unor astfel de soluţii, ar permite utilizarea sistemului de către elevi mai mici şi ar creşte utilizabilitatea software-ului.

MULŢUMIRI

Cercetările prezentate au fost susţinute de finanţarea Agence Nationale de la Recherche (DEVCOMP), de către proiectul FP7-REGPOT-2010-1 264207 ERRIC-Empowering Romanian Research on Intelligent Information Technologies şi de către proiectul POSDRU/107/1.5/S/76909 Valorificarea capitalului uman din cercetare prin burse doctorale (ValueDoc).

REFERINȚE

1. Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3, 4-5, 993–1022.
2. Budanitsky, A. and Hirst, G., 2006. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics 32, 1, 13–47.
3. Dascalu, M., Dessus, P., Trausan-Matu, S., Bianco, M., and Nardy, A., in press. ReaderBench, an Environment for Analyzing Text Complexity and Reading Strategies. In 16th Int. Conf. on Artificial Intelligence in Education (AIED 2013) Springer, Memphis, USA.
4. Dascalu, M., Dobre, C., Trausan-Matu, S., and Cristea, V., 2011. Beyond Traditional NLP: A Distributed Solution for Optimizing Chat Processing. In 10th International Symposium on Parallel and Distributed Computing (ISPDC 2011) IEEE, Cluj-Napoca, Romania, 133–138.
5. Dascalu, M., Trausan-Matu, S., and Dessus, P., 2012. Towards an integrated approach for evaluating textual complexity for learning purposes. In 11th Int. Conf. in Advances in Web-Based Learning (ICWL 2012), E. Popescu, R. Klamma, H. Leung and M. Specht Eds. Springer, Sinaia, Romania, 268–278.
6. Dascalu, M., Trausan-Matu, S., and Dessus, P., in press. Cohesion-based Analysis of CSCL Conversations: Holistic and Individual Perspectives. In 10th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2013) ISLS, Madison, USA.
7. Denhière, G., Lemaire, B., Bellissens, C., and Jhean-Larose, S., 2007. A semantic space for modeling children's semantic memory. In Handbook of Latent Semantic Analysis, T.K. Landauer, D.S. McNamara, S. Dennis and W. Kintsch Eds. Erlbaum, Mahwah, 143–165.
8. Galley, M. and McKeown, K., 2003. Improving Word Sense Disambiguation in Lexical Chaining. In 18th International Joint Conference on Artificial Intelligence (IJCAI'03), G. Gottlob and T. Walsh Eds. Morgan Kaufmann Publishers, Inc., Acapulco, Mexico, 1486–1488.
9. Geisser, S., 1993. Predictive inference: an introduction. Chapman and Hall, New York, NY.
10. Graesser, A.C., McNamara, D.S., and Kulikowich, J., 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. Educational Researcher 40, 5, 223–234.
11. Graesser, A.C., McNamara, D.S., Louwerse, M.M., and Cai, Z., 2004. Coh-Metrix: Analysis of text on cohesion and language. Behavioral Research Methods, Instruments, and Computers 36, 2, 193–202.
12. Graesser, A.C., Singer, M., and Trabasso, T., 1994. Constructing inferences during narrative text comprehension. Psychological Review 101, 3, 371–395.
13. Landauer, T.K. and Dumais, S.T., 1997. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review 104, 2, 211–240.
14. Landauer, T.K., Foltz, P.W., and Laham, D., 1998. An introduction to Latent Semantic Analysis. Discourse Processes 25, 2/3, 259–284.
15. Leacock, C. and Chodorow, M., 1998. Combining local context and WordNet similarity for wordsense identification. In WordNet: An electronic lexical database, C. Fellbaum Ed. MIT Press, Cambridge, MA, 265–283.
16. Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M., and Jurafsky, D., 2011. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In CONLL Shared Task '11 Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task ACL, Portland, OR, 28–34.
17. Manning, C.D., Raghavan, P., and Schütze, H., 2008. Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK.
18. Manning, C.D. and Schütze, H., 1999. Foundations of statistical Natural Language Processing. MIT Press, Cambridge, MA.
19. McCallum, A.K., 2002. MALLET: A Machine Learning for Language Toolkit.
20. McNamara, D.S., 2004. SERT: Self-Explanation Reading Training. Discourse Processes 38, 1–30.
21. McNamara, D.S., Boonthum, C., and Levinstein, I.B., 2007. Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In Handbook of Latent Semantic Analysis, T.K. Landauer, D.S. McNamara, S. Dennis and W. Kintsch Eds. Erlbaum, Mahwah, NJ, 227–241.
22. McNamara, D.S., Graesser, A.C., and Louwerse, M.M., in press. Sources of text difficulty: Across the ages and genres. In Assessing reading in the 21st century, J.P. Sabatini and E. Albro Eds. R&L Education, Lanham, MD.
23. McNamara, D.S., Louwerse, M.M., McCarthy, P.M., and Graesser, A.C., 2010. Coh-Metrix: Capturing linguistic features of cohesion. Discourse Processes 47, 4, 292–330.
24. McNamara, D.S. and Magliano, J.P., 2009. Self-explanation and metacognition. In Handbook of metacognition in education, J.D. Hacher, J. Dunlosky and A.C. Graesser Eds. Erlbaum, Mahwah, NJ, 60–81.
25. McNamara, D.S., O'Reilly, T.P., Rowe, M., Boonthum, C., and Levinstein, I.B., 2007. iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In Reading comprehension strategies: Theories, interventions, and technologies, D.S. McNamara Ed. Erlbaum, Mahwah, NJ, 397–420.
26. Miller, G.A., 1995. WordNet: A Lexical Database for English. Communications of the ACM 38, 11, 39–41.
27. Miltsakaki, E. and Kukich, K., 2000. The role of centering theory's rough-shift in the teaching and evaluation of writing skills. In 38th Annual Meeting on Association for Computational Linguistics ACL, Hong Kong, 408–415.
28. Nelson, J., Perfetti, C., Liben, D., and Liben, M., 2012. Measures of text difficulty: Testing their predictive value for grade levels and student performance. Council of Chief State School Officers.
29. Page, E., 1966. The imminence of grading essays by computer. Phi Delta Kappan 47, 238–243.
30. Sagot, B., 2008. WordNet Libre du Francais (WOLF) INRIA, Paris.

31. Tapiero, I., 2007. Situation models and levels of coherence. Erlbaum, Mahwah, NJ.
32. Trausan-Matu, S., Dascalu, M., and Dessus, P., 2012. Textual Complexity and Discourse Structure in Computer-Supported Collaborative Learning. In 11th Int. Conf. on Intelligent Tutoring Systems (ITS 2012), S.A. Cerri, W.J. Clancey, G. Papadourakis and K. Panourgia Eds. Springer, Chania, Grece, 352–357.
33. Trausan-Matu, S., Dascalu, M., and Rebedea, T., 2012. A system for the automatic analysis of Computer-Supported Collaborative Learning chats. In 12th IEEE Int. Conf. on Advanced Learning Technologies (ICALT 2012), C. Giovannella, D.G. Sampson and I. Aedo Eds. IEEE, Rome, Italy, 95–99.
34. van Dijk, T.A. and Kintsch, W., 1983. Strategies of discourse comprehension. Academic Press, New York, NY.
35. Wu, Z. and Palmer, M., 1994. Verb semantics and lexical selection. In 32nd Annual Meeting of the Association for Computational Linguistics, ACL '94 ACL, New Mexico, USA, 133–138.