

# Evaluarea învățării colaborative pe chat, pe baza analizei repetițiilor și altruismului

Ștefan Trăușan-Matu<sup>1,2,3</sup>, Roxana-Ioana Murăruș<sup>1</sup>

<sup>1</sup>Universitatea Politehnica din București  
Splaiul Independenței nr. 313, București,  
E-mail: stefan.trausan@cs.pub.ro, murarus.roxana@gmail.com

<sup>2</sup>Institutul de Cercetări în Inteligența Artificială  
Calea 13 Septembrie nr. 13, București

<sup>3</sup>Academia Oamenilor de Știință din România  
Splaiul Independenței nr. 54, București

**Rezumat.** Lucrarea prezintă un sistem de evaluare a studenților participanți la o sesiune de învățare colaborativă de tip chat pentru discuții în limba engleză. Fiecărui participant îi este asociat un concept diferit și studenții trebuie să dezbate polemic aceste concepte. Baza analizei este analiza folosirii de repetiții, punându-se accent pe funcțiile pe care acestea le pot îndeplini, de la socializare la sublinierea unei idei. Se folosesc tehnici de prelucrarea limbajului natural. Numărul și tipul de cuvinte repetate de fiecare participant în parte este o măsură a implicării studentului respectiv în fiecare fir de discuție. Evaluarea participanților constă în acordarea unui calificativ *Very good*, *Good*, *Satisfying* și *Unsatisfying* fiecăruia dintre ei pe baza numărului de cuvinte specific dezbaterilor, a celor din domeniul subiectelor de discuție, pe baza lungimii cuvintelor folosite și printr-o măsură a implicării lui în discuția referitoare la subiectul său și la subiectele celorlalți. Utilizatorul are posibilitatea de a seta procentul în care contează fiecare dintre aceste criterii și tematica asociată fiecărui participant.

**Cuvinte cheie:** polifonie, dialogism, repetiții, CSCL, suffix tree, altruism

## 1. Introducere

Datorită progresului webului și instrumentelor de colaborare care folosesc resursele acestuia, s-a ajuns la o nouă posibilitate de a învăța, adaptată webului social: învățare colaborativă sprijinită de calculator (Computer-Supported Collaborative Learning – CSCL). O modalitate foarte des folosită în CSCL este conversația pe sisteme de mesagerie instantanee (chat), în scopul de a dezbate și înțelege o anumită problemă (Stahl, 2006). Un exemplu, folosit în această lucrare, sunt temele date la cursul de Interacțiune Om-Calculator de la Departamentul de Calculatoare din Universitatea

Politehnica din București, la care studenții trebuie să dezbată pe chat, în echipe de patru participanți, tehnologiile colaborative timp de cel puțin o oră (Trăușan-Matu, Dascălu, Rebedea și Gartner, 2010).

O problemă în folosirea chat-urilor pentru CSCL este dificultatea și timpul mare necesitat pentru analiza acestora de către profesori în scopul furnizării de feedback și al notării studenților participanți (Trausan-Matu, 2010b). De aceea este foarte util să existe instrumente informatice de analiză a logurilor de chat, pentru aprecierea contribuțiilor participanților la procesul colaborativ de construire a cunoașterii. Sunt necesare însă tehnici de prelucrare a limbajului natural (NLP), fiind vorba de conversații în formă textuală.

Punctul de plecare în cercetările prezentate în lucrarea de față îl constituie ideea că între muzică și conversații, în general, dar și texte, în particular, există asemănări manifestate prin mai mulți factori (Trausan-Matu, 2013; Călinescu și Trausan-Matu, 2013). În acest sens a fost introdus modelul polifonic (Trăușan-Matu, Stahl și Sarmiento, 2007; Trăușan-Matu, 2010a; Trăușan-Matu, 2013; Trăușan-Matu, 2014; Trăușan-Matu et. al., 2014) și metoda polifonică de analiză a lor (Trăușan-Matu, 2013; Trăușan-Matu, 2014; Trăușan-Matu et. al., 2014), în care fenomenul repetiției are un loc principal (Tannen, 2009; Trausan-Matu, 2012). Acest model este direct legat de teoria dialogistică a lui Bahtin, care presupune că în orice fenomen lingvistic (în cazul nostru conversația), există mai multe voci, care se inter-animează, cu potențial de a forma structuri similare cu muzica polifonică (Bahtin 1970, 1981).

În această lucrare, spre deosebire de lucrările anterioare care consideră modelul polifonic, accentul este pus pe dimensiunea repetițiilor și a altruismului. Este prezentat și un sistem de analiză a chaturilor care este bazat pe analiza repetițiilor de cuvinte și sintagme. În cadrul acestuia, am considerat că evaluarea unui student se face pe baza numărului de cuvinte folosite specifice dezbaterilor, a celor specifice domeniului de discuție, a lungimii cuvintelor folosite și pe baza unui algoritm de evaluare a colaborării.

Corpusul de chaturi pentru testare este format din sesiuni de învățare prin colaborare dezvoltate de studenți din cadrul Facultății de Automatică și Calculatoare a Universității Politehnica din București. Subiectul de discuție este: Care dintre instrumentele de colaborare este cel mai util? Instrumentele alese pentru dezbateri sunt: chat forum, wiki și blog, fiecărui student

atribuindu-i-se unul din aceste subiecte (Trausan-Matu et al., 2010; Trausan-Matu, Dascalu și Rebedea, 2014).

Discuțiile și cuvintele din toate fișierele pe care le folosește aplicația sunt în limba engleză deoarece este o limbă de circulație mondială și există mai multe cercetări și statistici folosind sesiuni în limba engleză. Ar putea exista diferențe față de limba română, mai ales în ceea ce privește repetițiile, în sensul că s-ar putea repeta alte părți de vorbire. Alte diferențe legate de strategiile de repetiție în diferite limbi reprezintă încă un subiect de discuție asupra căruia se fac cercetări.

Lucrarea continuă cu o scurtă trecere în revistă a ideilor dialogismului și polifoniei, cu accent pe repetiții. Secțiunea a treia este dedicată repetițiilor iar a patra prezentării sistemului implementat.

## **2. Dialogism, repetiții, polifonie**

Conversațiile pe mesageria instantanee („chat”) cu cel puțin trei participanți facilitează dezvoltarea structurilor similare muzicii polifonice mai sus amintite (Trausan-Matu, Stahl și Sarmiento, 2007), spre deosebire de cele față în față, de cele la telefon sau de videoconferințe (de exemplu, skype). Această particularitate este generată de faptul că mai mulți participanți pot scrie replici în același timp, nu este un sigur participant care poate vorbi la un moment dat (care „ocupă podiumul” – „take the floor” în engleză). De exemplu, în conversații chat, de multe ori participanții folosesc simbolul '@' pentru a indica la ce participant sau replică se referă. Mediul de chat ConcertChat/VMT (Stahl, 2009), folosit și în cadrul cercetărilor noastre, oferă și facilitatea de a indica direct, printr-un clic, replica referită, prin aceste legături explicite încurajându-se și mai mult formarea de fire de discuție în paralel. Conversația se poate exporta, aceste legături explicite sunt înregistrate, se va putea face o analiză asupra acestora, inclusiv pentru identificarea firelor de discuție.

În afara legăturilor explicite, indicate prin referirea replicilor, există numeroase legături implicite, realizate prin perechi de adiacență, coreferințe, scheme retorice (Jurafsky și Martin, 2009), argumentări și, nu în cele din urmă, prin repetiții de cuvinte. Din punct de vedere practic, o diferență între cele două tipuri de legături este că cele explicite sunt, evident, direct accesibile, în timp ce cele implicite necesită pentru

identificare, de obicei, tehnici avansate de prelucrarea limbajului natural. O excepție sunt repetițiile de cuvinte sau sintagme. În plus față de faptul că sunt foarte ușor de identificat, acestea sunt și esențiale în conversații (Tannen, 2007).

Posibilitatea ca mai mulți participanți să scrie replici în același timp și referirea explicită duc și la posibilitatea apariției a mai multor fire de discuție în paralel (Trausan-Matu, Stahl și Sarmiento, 2007). Acest fenomen duce inevitabil la interferențe, la divergențe și convergențe care dau naștere la inter-animări (Trausan-Matu, Stahl și Sarmiento, 2007). Din perspectivă polifonică, fiecare fir de discuție poate fi considerat ca o voce, în sens generalizat (Trausan-Matu, Dascalu și Rebedea, 2014), care intră în relații contrapunctive cu celelalte. Cazul cel mai simplu de voce, în sensul pe care noi îl considerăm, este astfel o secvență de repetiții de cuvinte, care se pot transforma în artefacte (Trausan-Matu, 2012).

### 3. Forme ale repetiției

Repetiția este un fenomen omniprezent în conversații și, în general, în texte, având o importanță deosebită (Tannen, 2007). Pot fi identificate mai multe criterii în care putem clasifica formele repetiției în conversații. Se poate vorbi în primul rând, în funcție de ce categorii de cuvinte se repetă, de repetiții de:

1. cuvinte cheie (care corespund unor concepte) eventual stabilite sau impuse pentru a fi discutate în conversația chat, de exemplu, în cazul chat-urilor din CSCL;
2. cuvinte sau sintagme cheie din domeniul conversației – pentru detectarea apartenenței la domeniul considerat se pot folosi liste de cuvinte, ontologii sau metode cum ar fi LSA (Jurafsky și Martin, 2009) sau LDA (Blei, Ng și Jordan, 2003);
3. cuvinte sau sintagme cheie emergente – acestea sunt detectate ca urmare a apariției lor frecvente, prin numărare după eliminarea stopwords și stemming (pentru cuvinte cheie) sau folosind arbori de sufixe (pentru sintagme);
4. cuvinte sau sintagme cheie specifice dezbaterilor;
5. cuvinte sau sintagme cheie specifice schemelor retorice.

Primele trei categorii se referă la conținut, iar ultimele două la structuri discursive. În perspectiva polifonică, prima categorie trebuie să conțină voci (în sensul generalizat mai sus definit) în conversație, iar următoarea poate genera voci. A treia categorie conține, după părerea noastră, voci evidente. Ultimele două categorii pot fi indicatori ai patternurilor de inter-animare.

Repetițiile pot fi clasificate și după cel care le face. Dacă un participant face o repetare a ceva ce el a spus sau scris, putem spune că este o *reiterare*. Dacă un participant repetă un cuvânt sau sintagmă a unui alt participant, putem vorbi de o *preluare*. O diferență între cele două situații ar putea fi făcută și prin prisma altruismului, o preluare a unui cuvânt sau sintagmă de conținut indicând o posibilă atitudine altruistă (orientată dezinteresată către un altul). Există însă și varianta ca repetiția să fie pur și simplu un automatism repetitiv sau o polemică (dar și polemica poate fi altruistă).

Cuvintele pot fi repetate în forma lor inițială sau se pot folosi sinonime sau parafrazări, adică formularea de idei similare folosind cuvinte diferite. De asemenea, întrebări pot fi repetate prin transformare în afirmații, afirmații transformate în întrebări, repetiții ale unei replici cu un sigur cuvânt schimbat și repetiții având persoana sau timpul verbului modificat.

Repetițiile pot fi imediate sau îndepărtate temporal, în al doilea caz punându-se întrebarea dacă este într-adevăr o repetiție care constituie o legătură implicită a celor două apariții sau este o pură întâmplare, dacă este o reluare sau nu. De asemenea, există o ambiguitate în a determina dacă vorbim de o reluare a primei apariții considerate sau ambele apariții sunt reluări ale unei alte apariții anterioare.

O altă clasificare a repetițiilor este cea care consideră distribuția lor în timp, dacă sunt ritmice sau nu. Ritmul este o dimensiune importantă a discursului, mai ales în conversații (Tannen, 2007).

Este posibil ca o parte din repetițiile dintr-o conversație să fie automatisme, adică vorbitorii repetă ceea ce au spus alți vorbitori, reformulând sau reproducând exact, fără să se gândească, doar pentru a participa la dialog. Cercetările neurolingviștilor, așa cum specifică Killmer (2010) au arătat că persoanele bolnave de afazie (care suferă de imposibilitatea de a-și forma singuri cuvintele în vorbire în mod spontan) sunt capabile să repete cuvintele auzite. Acest lucru se poate explica prin faptul că reproducerea cuvintelor se face în altă parte a creierului decât cea folosită la generarea normală de limbaj. S-a observat că unele conversații

care au loc între persoane sănătoase seamănă cu o conversație cu o persoană ce suferă de afazie. În aceste condiții, este foarte greu de presupus cât din replicile oamenilor sănătoși reprezintă originalitate și cât automatism.

Continuăm cu o analiză a câtorva exemple de repetiții, cu o discuție privind funcțiile lor în conversație, plecând de la corpusul de chat-uri CSCL prezentat în introducere.

**Student1:** Hello all

**Student2:** Hello

Cele două replici pot fi considerate și un automatism cu rolul de socializare de la bun început creând premisele unui dialog. Este și cel mai ușor mod de a începe o discuție.

**Student1:** can you select a line?

**Student2:** you can select a line or parts of a line

**Student1:** you select with the mouse one line or word, it shows an arrow pointing to it. that is my reffer, right?

În prima parte, este un exemplu de transformare a unei întrebări în afirmație. Student1 reia asocierea de cuvinte în cea de-a doua replică a sa, pentru a evita o greșeală. S-ar putea asemana cu situația în care o persoană repetă un cod pentru a fi sigur că este cel auzit anterior.

**Student1:** i'm against blogs

**Student2:** why are you against blogs Student1?

Acesta este un exemplu invers celui anterior, în care o afirmație este transformată în întrebare.

**Student1:** please select a text before replying

... [câteva replici mai târziu]

**Student1:** Student2, select a text above before replying!

Prin repetarea întârziată de către aceeași persoană a unei replici, se încearcă sublinierea unei idei în contextul în care persoana vizată nu observă replica inițială. Motivul ar putea fi faptul că într-un chat replicile se succed foarte repede.

**Student1:** i have tons of very rellevant blogs

**Student2:** yes...but those are all over the net

**Student3:** but there are like ten tons more bad blogs

Prin repetarea cuvântului „but” se observă că Student2 și Student3 formează o voce comună împotriva ideii lui Student1. Cei doi aduc două argumente diferite împotriva aceleiași idei.

**Student1:** I will like to support this because I don't like them either, because most of them are a waste of space and time.

În acest exemplu se pune foarte mult accentul pe persoana I, adică părerea este una personală și este foarte important să se țină cont de acest lucru. Repetiția cuvântului „because” indică, ca și în exemplul de mai sus că sunt două argumente diferite aduse faptului că Student1 nu apreciază subiectul discuției.

**Student1:** you can save your chat room session

**Student2:** chat rooms are for communication

Repetiția cuvintelor chat room permite identificarea firului de discuție asociat instrumentului de colaborare respectiv.

#### **4. Aplicație de evaluare a participanților la un chat CSCL pe baza repetițiilor**

Aplicația dezvoltată se încadrează în seria celor bazate pe modelul polifonic (Trausan-Matu, Rebedea et. al, 2007; Trausan-Matu, Dascalu și Rebedea, 2014; Dascalu et al., 2013) și constă în analiza unei sesiuni de învățare prin colaborare sprijinită de calculator (CSCL) folosind chat și evaluarea studenților participanți. Originalitatea acestei aplicații constă în considerarea cuvintelor specifice dezbaterilor, a detectării repetițiilor de sintagme folosind abori de sufixe și considerarea aspectelor altruiste, prin analiza ratei referințelor explicite la topicile partenerilor față de cele proprii.

Aplicația este dezvoltată pentru a evalua orice tip de chat care urmează următorul șablon: există un număr de participanți, fiecare având asociată o temă de discuție, toate temele fiind din același domeniu. Un participant este pus să susțină un anumit subiect aducând argumente pro acestuia și argumente împotriva celorlalte subiecte.

Instrumentul folosit pentru discuțiile chat este VMT/ConcertChat (Stahl, 2009) deoarece permite utilizarea de referințe între replici, ceea ce ușurează dezambiguizarea atunci când apar mai multe fire de discuție, și evaluarea colaborării între participanți. ConcertChat permite salvarea conversației împreună cu referințele explicite în format html. Din acest format se poate genera un format XML conform schemei folosite la mai multe aplicații similare (Trausan-Matu et al., 2010) care este acceptat de aplicația dezvoltată.

Prima etapă de procesare constă în determinarea participanților la chat. Următorul pas este determinarea tuturor replicilor participanților. Se elimină replicile de tipul „Student1 joins the room” și „Student1 left the room” deoarece nu au relevanță în evaluare.

După determinarea replicilor pentru fiecare persoană în parte, se determină cuvintele folosite cel mai des de către toți studenții participanți deoarece sunt importante atât repetițiile făcute de aceeași persoană, cât și de persoane diferite. Fiecare din replici va fi separată în cuvinte. După eliminarea cuvintelor de tip „stop-words” (cuvinte care foarte frecvente și nu sunt de conținut, de exemplu, „și”, „sau”, „un”, „de”, „pentru” etc.) se vor contoriza cuvintele din chat repetate cel mai des în timpul discuției și numărul lor de apariții.

Repetițiile se consideră a nu fi case-sensitive, adică se vor considera repetiții toate cuvintele care conțin aceleași caractere în aceeași ordine, indiferent dacă acestea sunt litere mari sau mici. Se iau în considerare și repetițiile imediate și cele făcute la una sau mai multe replici distanță.

Pornind de la conținutul tuturor replicilor persoanelor se creează un arbore care va cuprinde toate sufixele posibile, un „Suffix Tree” (<http://www.allisons.org/ll/AlgDS/Tree/Suffix/>), care permite identificarea repetițiilor de sintagme. Plecând de la această structură se determină și toate perechile de cuvinte care se repetă.

O problemă importantă în contextul unei sesiuni de învățare prin colaborare este analiza participării fiecărui student la discuție. Această analiză poate porni de la cât de mult a influențat studentul conversația. În sesiunile de învățare colaborativă se pornește de la o anumită temă, propusă de un profesor sau de participanți. Este foarte important și logic să se folosească pe parcursul conversației cât mai multe cuvinte specifice acelei teme. În cadrul evaluării se va urmări numărul de astfel de cuvinte utilizate de către fiecare participant.

Pe de altă parte, prezența multor legături explicite sau implicite între replici ar putea să sugereze că s-a ajuns la o convergență, adică la găsirea unei soluții într-o sesiune de dezbatere a unei probleme, că s-a ajuns la un punct comun. Pe de altă parte se poate, de asemenea, să apară un fir de discuție care nu are nimic în comun cu cele dinainte și care se poate spune că se află pe o poziție divergentă.

Pentru fiecare student participant se va da un calificativ și, de asemenea, mai multe detalii legate de modul în care s-a obținut acel



calificativ. La evaluarea sudeților vor contribui numărul de cuvinte din lista subiectelor discutate, numărul de cuvinte specifice dezbaterilor, lungimea cuvintelor utilizate și implicarea în discuție. Procentele în care vor conta fiecare dintre acestea la nota finală și pe baza cărora se vor determina calificativele pot fi modificate de utilizator.

În evaluarea fiecărui participant în parte se pleacă de la lista conținând cuvintele pentru fiecare participant din care au fost eliminate stop words. Evaluarea se face pe baza a 4 criterii:

1) **Repetiția de cuvinte specifice domeniului de discutat pe parcursul sesiunii.** Aceste cuvinte sunt înregistrate într-un fișier și sunt concepute din domeniul din care fac parte și subiectele asociate înainte de începerea discuțiilor. Acest fișier trebuie modificat în funcție de tematica discuției. Corpusul considerat pentru analiză la testare a conținut chat-uri în care participanții trebuiau să susțină folosirea unuia dintre următoarele instrumente: chat wiki, blog, forum. Așadar, ontologia considerată este cea a sistemelor de colaborare pe web.

2) **Repetiția de cuvinte specifice dezbaterilor.** Acesta este un criteriu original propus de această lucrare. Aceste cuvinte sunt citite dintr-un fișier dedicat. Cuvintele de acest tip pot susține o idee (ex: „agree”, „true”, „certain”, „exactly” etc.), pot fi precursori unui argument (ex: „because”, „but”), cuvinte care solicită noi informații („why”, „how” etc.) sau pot avea alte funcții. Pentru fiecare participant se determină numărul de cuvinte de acest tip. Dacă există un număr mare de astfel de cuvinte, se presupune că participantul care le-a emis a avut replici în care și-a exprimat acordul, dezacordul sau a pus întrebări. Toate acestea presupun că participantul a contribuit consistent la dezbateri.

3) **Lungimea cuvintelor** Se va determina, pentru fiecare student numărul de cuvinte de dimensiune mai mare decât un anumit număr (am considerat 5).

4) **Implicarea altruistă în discuție.** Evaluarea pe baza acestui criteriu se face prin analiza gradului în care un participant se referă atât la topicul asociat lui cât și la al celorlalți (Trausan-Matu, 2014). Folosind algoritmul

prezentat mai jos se determină numărul de replici în care vorbitorul folosește cuvinte asociate subiectului său. De asemenea, și aici intervine partea de colaborare, se determină numărul de replici în care vorbitorul face referire la alte replici, în care se discută despre subiectele celorlalți participanți la discuție.

```

for each r in replici repeat
    u.iv = participant(u).subiect
    u.rv = (u.referinta_explicita).cv
if u.iv is in u.cv then
    nr(iv=cv) = 1+nr(iv=cv)
    for each i in u.rv repeat
        if i <> iv then nr(iv<>rv) = 1+nr(iv<>rv)

```

Variabilele folosite reprezintă (în Figura 1 este ilustrat un fragment de chat cu indicarea iv, cv, rv, a numărului replicii și a replicilor referite; F,C,W,B reprezintă „forum”, „chat”, „wiki” și „blog”):

- u, replica;
- iv, subiectul asociat fiecărui participant;
- cv, subiectele din replică;
- rv, subiectele din replica referită explicit prin facilitatea oferită de ConcertChat;
- nr(iv=cv), o estimare a numărului de replici în care un participant vorbește despre subiectul asociat lui;
- nr(iv<>rv), o estimare a numărului de cazuri în care un participant a făcut referire explicit la o replică în care se vorbește despre un subiect care nu-i aparține.
- Fiecare din cele patru criterii mai sus prezentate contează cu un anumit procent în determinarea notei finale. Procentele pot avea valorile setate implicit („default”) din program sau pot fi introduse de către utilizator din interfața grafică. Evaluarea se face folosind distribuția normală, după modelul clopotului lui Gauss deoarece aceasta se folosește atunci când variabilele tind să se găsească în jurul mijlocului intervalului de valori. Astfel, numărul de persoane care vor avea calificativul *Very good* va fi aproximativ același cu numărul de persoane care vor avea

calificativul *Unsatisfying*. Cei mai mulți studenți vor avea calificativele *Good* și *Satisfying*, distribuția între acestea făcându-se în mod egal.

Nr	Student	Utterance	iv	cv	rv	Ref
74	bogdan	if you r willing to spend a lot of time, that may be true for a small wiki , but when you a trying to develop an encyclopedia ....you have absolutely no chance	C	W		70
75	Raluca	Messages posted to a forum are publicly available for some time, which is uncommon in chat rooms, with only few exceptions.	F	F	W	74
76	florin	why are blogs search-engine friendly? I think wikis are more search-engine frierly! :P		C		
			W	B	B	73
77	bogdan	chats work as long as you have an internet connection. ...	C	C	C	72
78	elena	because you can use keywords	B		B	p 71
79	florin	never say never - what about Wikipedia ?!	W		W	74
80	florin	can you detail on that feature a bit?	W			78
81	Raluca	in terms of 'coding' your application, I bet yours is the hardest	F			77
82	Raluca	and only that could take up a lot of time	F		C	77
83	bogdan	it may be, but you're only going to do it once....and then use it like that for a very long time	C			81
84	elena	the blogs can be grouped by interest and all the articles can be full of keywors hat are search engine friendly	B		B	p 81
85	elena	do not forget that a blog is the best way to promote a site	B		B	
86	Raluca	same with forums, and reading the posts looks easier	F	F	B	84
87	florin	ok, thanks for the info	W		B	84
88	bogdan	sorry , but the est way is a message to all your friends. using a CHAT	C	C	B	85
89	Raluca	but if you want to study, you have a lot of information on the forums, moreover, if you have questions people can take their time to answer you	F	F	C	88

Figura 1. Fragment de chat

Am considerat că toate persoanele care au nota 0 primesc calificativul *Unsatisfying*. Dacă nu există astfel de persoane, ultimii 10% dintre participanți vor primi acest calificativ. Același procent va primi calificativul *Very good*, de acest lucru având parte cei cu notele cele mai mari. Restul de persoane vor fi împărțite în mod egal: primii 40 % care urmează după cei 10% având notele cele mai mari, vor primi calificativul *Good*, iar restul *Satisfying*. Dacă există un număr impar de persoane pentru cei 80 % rămași după primul pas vor fi mai mulți cei care vor avea calificativul *Good*.

Utilizatorul are posibilitatea de a modifica procentele pentru fiecare din criteriile de evaluare: cuvintele din lista de dezbateri, cuvintele specifice domeniului discutat, lungimea cuvintelor folosite și implicarea în discuție dată de algoritmul de colaborare. Există valori implicite („default”) care pot fi selectate bifând checkboxul cu eticheta „Default”.

Apăsând butonul „OK” vor apărea rezultatele în panelul de jos în fereastra principală, astfel: în panelul din stânga se vor vedea cuvintele repetate și numărul lor de apariție, iar în panelul din stânga, în fereastra în care se poate face scroll, vor apărea numărul de cuvinte de dezbateri, cele specifice domeniului și calificativul pentru fiecare dintre studenți, așa cum se poate vedea în Figura 2. O dată cu apariția rezultatelor, va fi vizibil un

nou buton, cu eticheta „More on results”. Apăsându-l, va apărea o nouă fereastră („Detailed results”), cu informații mai multe despre fiecare student: cât de mult s-a implicat în discuția privind subiectul său și privind subiectele celorlalți. De asemenea, va apărea un tabel cu cele mai repetate 20 de cuvinte de către fiecare dintre ei. Tot în fereastra de detalii, în partea de jos se vor afișa secvențele de mai mult de 3 cuvinte care se repetă în chat de către toți participanții. Rezultatele sunt salvate și într-un fișier, pentru eventuale consultări ulterioare.

Pentru implementare a fost folosit limbajul de programare Python. Pentru interfață, s-a folosit „pygtk”, un wrapper peste biblioteca GTK+ 4. Deși este dezvoltată pe sistemul de operare Windows 7, aplicația este independentă de platformă, adică funcționează pe sistemele de operare Windows și Linux, fiind testată pe amândouă.

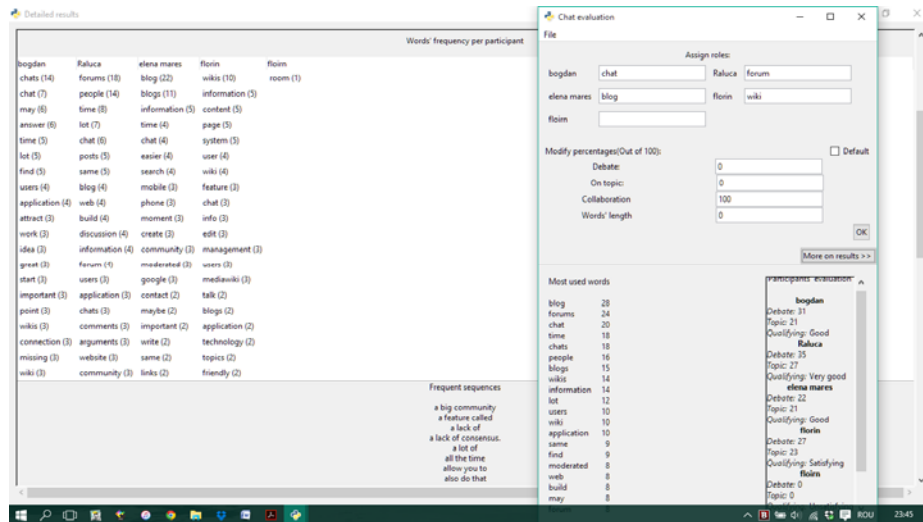


Figura 2: Interfața sistemului dezvoltat

Pentru testare, am analizat mai multe chat-uri, modificând procente în care contează fiecare criteriu folosit în calculul notelor. Subiectele asociate participanților sunt cele care reies din discuțiile acestora (inițial, fiecare student se prezintă și își alege subiectul de susținut).

## 5. Concluzii și dezvoltări ulterioare

Ideile implementate în sistemul prezentat se pot dezvolta ca o metodă de evaluare chiar total automată, care nu implică subiectivitate. Se deschide însă o discuție lungă privind cât de bine ar fi ca o notă să fie acordată automat, fără nici un fel de factor uman. Evaluatorii umani nu sunt perfecți, se invocă de multe ori subiectivitatea unui evaluator. Automatizarea evaluării are însă și ea neajunsurile ei: este foarte greu să se dezvolte o metodă de prelucrare a informațiilor comparativă cu modul de selectare a informației relevante de către un om. Ar fi necesar un program care să înțeleagă limbajul natural, deziderat încă departe de a fi obținut.

Metoda de evaluare ar putea fi modificată, pentru a se lua în considerare mai multe criterii. S-ar putea ține cont de coreferințe, care se fac atunci când mai multe replici sau expresii se referă la o aceeași replică anterioară. De asemenea, se pot lua în considerare referințele implicite (aluziile) sau s-ar putea acorda o notă bazată pe funcțiile repetițiilor identificate în replicile fiecărui participant. Bineînțeles, ar fi dificil de determinat funcțiile, dar într-o primă fază s-ar putea forma o bază de date cu șabloane de repetiții și funcția asociată lor. Referitor la repetițiile de secvențe de cuvinte, s-ar putea determina frecvența lor de apariție, pornind tot de la structura de tip Suffix Tree. S-ar putea, de asemenea, filtra în funcție de anumite criterii (cuvinte din secvențe care fac parte din diverse ontologii).

## Referințe

- Bahtin, M. (1970), *Problemele poeziei lui Dostoievski*. București: Ed. Univers.
- Bakhtin, M. M. (1981), *The Dialogic Imagination: Four Essays*. Austin: University of Texas Press.
- Blei, D., Ng, A. and Jordan, M. (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022
- Călinescu, A., Trăușan-Matu, Ș. (2013), A System For Sonification Of Chat Conversations, *Annals of the Academy of Romanian Scientists Series on Science and Technology of Information*, vol.6, nr.2, 2013, pp.23-42.
- Dascalu, M. Trausan-Matu, S., Dessus P. (2013), Voices' Inter-Animation Detection with ReaderBench Modelling and Assessing Polyphony in CSCL Chats as Voice Synergy, *2nd IEEE International Conference on Systems and Computer Science (ICSCS)*, Villeneuve d'Ascq, France, pp. 281-285

- Jurafsky, D., Martin, J.H. (2009), *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Pearson Prentice Hall.
- Killmer, H., *Doubles - A study of repetition in the conversation of persons with aphasia*, PhD Thesis, 2010.
- Stahl, G. (2006), *Group Cognition: Computer Support for Building Collaborative Knowledge*, Cambridge MA: MIT Press.
- Stahl, G. (Ed.), (2009). *Studying Virtual Math Teams* Springer, Boston
- Tannen, D., *Talking Voices* (2007). Repetition, Dialogue and Imagery in Conversational Discourse. Cambridge University Press, 2nd edition.
- Trăușan-Matu, Ș. (2010a), The Polyphonic Model of Hybrid and Collaborative Learning, In Wang, F.,L., Fong, J., Kwan, R.C., *Handbook of Research on Hybrid Learning Models: Advanced Tools Technologies and Applications*, Information Science Publishing, Hershey, New York, pp 466-486
- Trăușan-Matu, S. (2010b). Automatic Support for the Analysis of Online Collaborative Learning Chat Conversations. In P. Tsang, S. K. S. Cheung, V. S. K. Lee, & R. Huang (Eds.), *Hybrid Learning* (Vol. 6248, pp. 383–394). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Trăușan-Matu, S.(2012), Repetition as Artifact Generation in Polyphonic CSCL Chats, *Third International Conference on Emerging Intelligent Data and Web Technologies*, IEEE Conference Publications, pp. 194-198
- Trăușan-Matu, Ș. (2013). Studiu asupra perspectivei polifonice în analiza discursului în texte, în conversații și în muzică. Raport de cercetare ICIA
- Trăușan-Matu, Ș. (2014), Polyphonic Design, Conduct Experience, and Evaluation in CSCL Chats, *Annals of the Academy of Romanian Scientists Series on Science and Technology of Information*, Volume 7, Number 2/2014, pp. 21-34.
- Trăușan-Matu, Ș., Dascălu, M., Rebedea, T., Gartner, A. (2010) Corpus de conversatii multi-participant si editor pentru adnotarea lui, *Revista Romană de Interactiune Om-Calculator*, Vol.3, No.1, pp. 53-64
- Trăușan-Matu, Ș., Dascălu, M., Rebedea, T., (2014), PolyCAFe—automatic support for the polyphonic analysis of CSCL chats *International Journal of Computer-Supported Collaborative Learning*, 06/2014; Volume 9(2), Springer, pp. 127-156, 2014
- Trăușan-Matu, Ș., Rebedea, Ș., Dragan, A., Alexandru, C. (2007), Visualisation of Learners Contributions in Chat Conversations , in Fong, J., Wang, P. (Eds.), *Blended Learning*, Pearson Prentice Hall, pp. 215-224.
- Trăușan-Matu, Ș., Stahl, G., & Sarmiento, J. (2007), Supporting Polyphonic Collaborative Learning. *E-service Journal*, 6, pp. 58-74