

# Identificarea actelor de vorbire în dialogurile purtate pe chat\*

Ștefan Trăușan-Matu<sup>1,2</sup>, Costin Chiru<sup>1</sup>, Radu Bogdan<sup>1</sup>

<sup>1</sup>Universitatea Politehnica București

<sup>2</sup>Institutul de Cercetări în Inteligența Artificială al Academiei Române

*E-mail: trausan@cs.pub.ro*

**Rezumat.** Lucrarea prezintă două abordări de analiză a dialogurilor purtate în cadrul unui chat, în vederea identificării actelor de vorbire. Amândouă adnotează semiautomat un corpus de dialog, după care, în cea de-a doua se folosește un model Marlov ascuns pentru a învăța probabilități de succesiune a actelor de vorbire. Acest model este apoi folosit pentru identificarea actelor de vorbire în noi dialoguri.

**Cuvinte cheie:** acte de vorbire, DAMSL, modele Markov ascunse.

## 1. INTRODUCERE

Înțelegerea limbajului natural din transcrierea dialogurilor ridică probleme specifice față de textele obișnuite. Dialogurile au o structură asemănătoare unui joc, cu mutări sau replici, (în engleză, “utterance”), fiecărui participant venindu-i rândul să spună ceva, să joace (în engleză, “turn taking”). Mutările pot fi clasificate în anumite categorii, denumite acte de vorbire (în engleză, “speech acts”) a căror succesiune nu este întâmplătoare [2,3,4].

Lucrarea de față prezintă două abordări, una care folosește euristici de adnotare a actelor de vorbire și o a doua care învață succesiuni posibile de acte de vorbire din sesiuni de chat adnotate semiautomat în limbajul DAMSL (Dialog Act Markup in Several Layers [1,4]). În cea de-a doua abordare se folosesc pentru învățare și apoi recunoaștere modele Markov ascunse [4,5]. Adnotarea semiautomată inițială folosește euristici de adnotare care consideră anumite secvențe de cuvinte (în engleză, “cue phrases” [4]), la care au fost adăugate și adnotări manuale. Modelul Markov învățat este apoi folosit pentru adnotarea de dialoguri neadnotate [4].

Corpusul de dialog este constituit din mai multe sesiuni de „chat” în care se discută rezolvarea unor probleme de matematică. Aceste sesiuni au fost realizate în

---

\* A apărut într-o versiune puțin modificată în Ștefan Trăușan-Matu, Costin Pribeanu (Eds.), Interacțiune Om-Calculator 2004, ISBN 973-718-053-4, Editura Printech, București, 2004, pp. 206-214.

cadru proiectului Virtual Math Teams (vezi, <http://mathforum.org/wiki/VMT/>), condus de prof. Gerry Stahl de la Universitatea Drexel din Philadelphia, SUA. Acest proiect își propune să dezvolte instrumente informatice pentru asistarea rezolvării distribuite de probleme.

Lucrarea conține în continuare o prezentare a problematicii actelor de vorbire și a limbajului de adnotare DAMSL. Secțiunile 3 și 4 prezintă cele două abordări. Lucrarea este încheiată de concluzii și direcții de continuare.

## 2. ADNOTAREA ACTELOR DE VORBIRE

Teoria actelor de vorbire a fost introdusă de Austin și dezvoltată de Searle [2,3]. Actele de vorbire au fost clasificate în diferite feluri. De exemplu, J. Austin face o distincție între actele *constatative*, care descriu, relatează, consemnează o stare de fapt și *performative*, care presupun efectuarea unei acțiuni sau a unei schimbări a stării de fapt. Cele din urmă folosesc verbe de genul “a ordona”, “a promite”, “a jura” etc., și li se poate evalua rezultatul prin reușită sau eșec. Plecând de la verbele performative (din limba engleză), Austin distinge cinci clase de acte ilocutionare [2]:

- *verdictive*, care dau un verdict, folosesc “a considera”, “a estima”;
- *exercitive*, decizionale, folosesc “a ordona”, “a cere”;
- *comisive (promisive)*, care exprimă o angajare (“a garanta”, “a plănuși”);
- *comportamentative*, exprimă o atitudine, un comportament (“a critica”, “a mulțumi”);
- *expoziție*, (“a obiecta”, “a răspunde”).

John R. Searle identifică mai multe tipuri de acte de vorbire:

- *ilustrative* (“representatives”), care reprezintă o stare de fapt: aserțiuni, descrieri;
- *comisive* (“commissives”), care implica vorbitorul într-un curs viitor de acțiuni: promisiuni, amenințări;
- *directive* ce atrag atenția asupra efectuării unei acțiuni: comenzi, cereri;
- *declarații*, care aduc o anumită stare de lucruri: denumire, arestare, casatorie, binecuvântare;
- *expresive*, care indica starea psihologica sau atitudinea mentală a vorbitorului: felicitări, mulțumiri, scuze;
- *verdicative*, care dau o apreciere: judecată, iertare.

O încercare recentă de a dezvolta schemele de adnotarea actelor de vorbire este arhitectura DAMSL care codifică pe diferite nivele informațiile referitoare la replicile unui dialog. Două dintre aceste nivele, *funcția de anticipare* (FLF, în engleză, “forward looking function”) și *funcția de adaptare regresivă* (BLF în engleză, “backward looking function”) sunt extensii ale actelor de vorbire ce extrag

notiuni legate de structura dialogului cum ar fi perechile de adiacenta sau notiunile de grupare.

Funcția de anticipare a unei notiuni corespunzând unor acte de vorbire de genul Searle/Austin e bazată pe tipul actelor ce pot apărea într-un dialog [1,4]:

<b>Afirmații</b>	o afirmație făcută de cel ce vorbește
<b>Cerere de informație</b>	întrebare pusă de cel ce vorbește
<b>Confirmare</b>	întrebare pentru confirmarea unor informații
<b>Influența asupra ascultătorului</b>	<b>= directivele lui Searl</b>
<b>Opțiuni</b>	o sugestie slabă sau o listă de opțiuni
<b>Ordin</b>	este de fapt o comandă
<b>Influența asupra vorbitorului</b>	<b>= directivele lui Austin</b>
<b>Ofertă</b>	vorbitorul se oferă să facă ceva și așteaptă confirmarea
<b>Execuție</b>	vorbitorul face ceva
<b>Acte Convenționale</b>	
<b>Deschidere</b>	mesaje de întâmpinare
<b>Incheiere</b>	mesaje de închidere
<b>Mulțumire</b>	mulțumiri și răspunsuri la mulțumiri

Funcția de adaptare regresivă se bazează pe relația dintre notiunea curentă și alte notiuni care au fost enunțate înainte de către alte persoane. Aceasta include acceptarea/refuzarea propunerilor precum și notiuni de grupare [1,4]:

<b>Acord</b>	răspunsul vorbitorului la o propunere anterioară
<b>Acceptă</b>	acceptarea propunerii
<b>Acceptă Parțial</b>	acceptă o parte a propunerii
<b>Poate</b>	nici nu acceptă, dar nici nu refuză
<b>Refuză</b>	refuzarea propunerii
<b>Refuză Parțial</b>	refuzarea parțială a propunerii
<b>Abținere</b>	amană/evită răspunsul
<b>Răspuns</b>	răspunsul la o întrebare

<b>Înțelegere</b>	dacă vorbitorul înțelege noțiunea anterioară
<b>Semnal de neînțelegere</b>	vorbitorul nu a înțeles
<b>Semnal de înțelegere</b>	vorbitorul a înțeles
<b>Aprobare</b>	demonstrare prin continuare
<b>Repetare/Reformulare</b>	demonstrare prin repetiție/reformulare
<b>Completare</b>	demonstrare prin completare prin colaborare

### 3. EURISTICI DE ADNOTARE SEMIAUTOMATĂ

Algoritmul folosit realizează recunoașterea actelor de vorbire dintr-o discuție de tip chat. S-au grupat actele de vorbire în categoriile FLF și BLF (conform modelului DAMSL). Recunoașterea actelor de vorbire se bazează pe recunoașterea verbelor din propoziție (folosind WordNet [6]), precum și a unor expresii standard (“cue phrases”) pentru anumite acte de vorbire.

Deoarece programul de recunoaștere este conceput pentru o discuție de tip chat (cu replici scurte), s-a tratat fiecare replică în funcție de primul verb găsit în propoziție. Verbele sunt grupate de asemenea în două categorii: “auxiliare” (be, have, do...) și “neauxiliare” (orice alt verb).

Tipurile de acte de vorbire recunoscute de algoritm și condițiile care trebuie îndeplinite pentru fiecare tip sunt:

#### **Funcția de anticipare**

a) *Statement* (afirmație) – trebuie să îndeplinească una din următoarele condiții:

- primul verb din propoziție (auxiliar sau nu) să fie precedat de un pronume, substantiv sau adverb.

Ex: I hope we can solve this.  
This problem is difficult.  
I already done that.

- propoziția să conțină o formă prescurtată, de gen pronume + verb auxiliar (pentru aceste prescurtări s-a creat o bază de date locală, deoarece nu sunt recunoscute de WordNet).

Ex: I'm new at that.  
There's one variable.  
That's why we are doing this.

- orice altă propoziție care conține un verb și nu a putut fi încadrată în nici o categorie de acte de vorbire.

Ex: Tried that.  
Checking...

b) *Info\_request* (întrebare) – trebuie să îndeplinească una din următoarele condiții:

- propoziția să se termine cu semn de întrebare.

Ex: Any other replies?

- propoziția să înceapă cu un verb auxiliar urmat de un pronume sau substantiv.

Ex: Did you send the picture

Can somebody help me

- propoziția să înceapă cu un adverb specific (where, when etc.) sau cu o prescurtare de gen adverb + verb auxiliar.

Ex: How long is the segment  
Where's the picture

c) *Action directive* (comandă/rugaminte) – trebuie să îndeplinească una din următoarele condiții:

- propoziția să conțină o expresie specifică (let's, please etc).

Ex: Let's try to solve the problem.  
Please make sure everyone understands the answer.

- propoziția să înceapă cu un verb (auxiliar sau nu) aflat în forma de baza (la infinitiv).

Ex: Feel free to ask about anything that seems unclear.  
Check this out. - *Action directive*  
Checking... - *Statement*

- propoziția să înceapă cu "don't" urmat de un verb.

Ex: Don't try to calculate.  
Don't be sad.

d) *Conventional* (formule de salut, mulțumiri, scuze etc.)

- propoziția să conțină o expresie specifică (hi, hello, bye, thanks, sorry etc.)

e) *Other forward function* – orice propoziție care nu a putut fi încadrată în nici o altă categorie de acte de vorbire.

**Funcția de adaptare regresivă**

a) Agreement (Accept, Maybe, Reject)

- propoziția să conțină o expresie specifică pentru fiecare tip de acord (agreement) și să nu existe nici un verb în propoziție, înainte de expresia respectivă.

Ex: Yes, there is. - *Accept, Statement*  
He said yes. - *Statement*

b) Understanding (Acknowledge)

- propoziția să conțină o expresie specifică (ok, okay, alright etc.) și să nu existe nici un verb în propoziție, înainte de expresia respectivă.

Ex: OK guys, I have to go. - *Acknowledge, Statement*  
I feel alright. - *Statement*

c) *Answer* (răspuns)

- replica să fie de tip *agreement* și să fie precedată de o replica tip *info\_request* a altui user (caut un *info\_request* in ultimile 10 replici).

Ex1: Are you all there? - *Info\_request*  
Yes - *Answer*  
Ex2: That means two equations. - *Statement*  
Yes - *Accept*

- replica să fie de tip *statement*, *action\_directive* sau *other\_forward\_function* și să fie precedată direct de un *info\_request* al altui utilizator (verific doar ultima replică).

Ex:       Where's the picture               - *Info\_request*  
          I can't get it                       - *Statement, Answer*

#### 4. IDENTIFICAREA ACTELOR DE VORBIRE PE BAZA UNUI MODEL MARKOV ASCUNS

A doua abordare folosește pentru recunoașterea actelor de vorbire modelul canalului comunicațional cu perturbații introdus de teoria informației a lui Shannon [4,5]. Se consideră că avem un mesaj inițial, format din acte de vorbire, ce intră în canal și textul dialogului care este ieșirea din acesta. În aceste condiții, pentru a determina efectul perturbațiilor din canal, se vor calcula (pentru fiecare semnal în parte) probabilitățile de a fi un anumit semnal, iar după aceea se va lua în considerare semnalul a cărei probabilitate este maximă.

În cazul identificării actelor de vorbire avem mai multe semnale care pornesc de la sursă și mai multe semnale care ajung la destinație, iar în acest caz se pune întrebarea: "Care este cea mai probabilă secvență ce a pornit de la sursă și care a ajuns la destinație în forma dată?"

Fie  $D=d_1d_2d_3\dots d_n$  secvența de semnale care porneste de la sursa și  $E=e_1e_2e_3\dots e_n$  secvența de semnale care ajunge la destinație. Se pune problema calculării unei secvențe  $D^*$ , care maximizează probabilitatea ca la ieșire să se obțină secvența  $E$  condiționată de intrări:

$$D^* = \operatorname{argmax} P(D|E) = \operatorname{argmax} ((P(E|D) * P(D)) / P(E))$$

Cum ieșirea este întotdeauna aceeași, înseamnă că  $P(E)$  este constantă și cum pe noi ne interesează de fapt secvență maximă și nu valoarea maximă, nu vom mai ține cont de  $P(E)$ . Ne interesează, prin urmare, maximul produsului dintre probabilitatea  $P(E|D) * P(D)$ , adică:

$$D^* = \operatorname{argmax} (P(E|D) * P(D))$$

Dacă expandăm  $P(D)$  obținem:

$$P(D) = P(d_1) * P(d_2|d_1) * \dots * P(d_n|d_1, d_2, \dots)$$

Această probabilitate este greu de calculat și atunci ea se aproximează considerând ipoteza unui lanț (model) Markov de ordin  $k$ , adică faptul că fiecare semnal în parte depinde doar de  $k$  semnale anterioare, cu valori tipice  $k=1$  (când avem practic un automat finit cu tranziții probabiliste, caz de multe ori denumit model Markov fără a mai preciza un ordin) sau  $k=2$ . Pentru  $k=1$  se va calcula  $P(D)$  că:

$$P(D)=P(d_1)*P(d_2|d_1)*P(d_3|d_2)...*P(d_n|d_{n-1})$$

unde  $P(d_i)$  este probabilitatea de apariție a fiecărui tip de semnal în parte și  $P(d_i|d_{i-1})$  este probabilitatea condiționată să avem  $d_i$  dacă am avut înainte  $d_{i-1}$ , altfel spus, să avem bigrame. În concluzie trebuie să calculăm probabilitățile bigramelor. Dacă se consideră  $k=2$ , avem o formulă similară în care intervine  $P(d_i|d_{i-1},d_{i-2})$ , adică avem nevoie de probabilitățile trigramelor.

Pentru  $P(E|D)$  se face o simplificare similară:

$$P(E|D)=P(e_1|d_1)*P(e_2|d_2)...*P(e_n|d_n)$$

Pentru calculul bigramelor s-a procedat în următorul mod: Inițial s-au numărat toate tagurile din text: `nr_total`. După aceea, s-au numărat tagurile din fiecare tip (`nr_act_vorbire`) în parte: `nr_accept`, `nr_reject`, `nr_maybe`, `nr_accept_part`, `nr_reject_part`, `nr_none`. În continuare s-au calculat probabilitățile de apariție ale fiecărui tip ca fiind:

$$\text{prob\_act\_vorbire}=\text{nr\_act\_vorbire}/\text{nr\_total}$$

Valorile obținute pentru un corpus de chat și un subset de acte de vorbire au fost:

Act de vorbire	secvențe tipice de cuvinte	Număr apariții	Probabilitate
ACCEPTĂ PARȚIAL	ok, but; k, but; yeap, but; yes, but	8	0.003579
REFUZĂ PARȚIAL	not really	1	4.47427E-4
ACCEPTĂ	k, ok, yeap, yes, yup, right, yeah, ya, okay, true, exactly, alright, probably	391	0.174944
REFUZĂ	nope, no, nop, don't want, disagree, don't think, nah	71	0.031767
POATE	maybe, don't know, guess, whatever	27	0.012081
Neadnotat		1737	0.777181

Tabelul de mai sus este folosit și pentru calculul probabilităților  $P(e_i|d_i)$ , unde  $e_i$  sunt secvențe tipice de cuvinte.

Pasul următor a fost numărarea succesiunilor `nr_Act1_Act2`, adică `nr_AP_AP`; `nr_AP_RP`; `nr_AP_A`; `nr_AP_R`; ș.a.m.d. Ultimul pas a fost calculul probabilităților :

prob\_Act1\_Act2=nr\_Act1\_Act2/nr\_Act1

Valorile obținute au fost:

probabilitate	AP	RP	A	R	P	N
AP	0	0	0.25	0	0	0.75
RP	0	0	0	0	0	1
A	0.00767	0	0.25575	0.01279	0.01279	0.711
R	0	0	0.21127	0.15493	0	0.6338
P	0	0	0.22222	0.03704	0.03704	0.7037
N	0.00288	0.00058	0.15429	0.03109	0.01209	0.7985

Similar s-a procedat pentru calculul trigramelor. Deoarece multe valori sunt 0, a fost făcută o netezire folosind metoda lui Katz [4].

Tabelele cu probabilități ale actelor de vorbire, a bigramelor și trigramelor pot fi folosite ca model pentru adnotarea unui dialog nou, pe baza alegerii argumentului care maximizează formula anterioară  $D^* = \text{argmax}(P(E|D) * P(D))$ , pentru un caz particular dat. În acest scop a fost folosit algoritmul Viterbi, bazat pe o schemă de programare dinamică [4,5].

Mai jos este arătat un fragment din dialogul adnotat în DAMSL, în care atributul "agreement" a fost pus automat de sistem:

```
<Turn Id="T272" Speaker="AvrilLR" time="8:52:46 PM">
  <Utt Id="Utt272" Info-level="" Conventional="" Time="8:52:46 PM"
    Influence-on-listener="" Influence-on-speaker="" Agreement="None"
    Answer="">I just ignored the other number</Utt>
</Turn>
<Turn Id="T273" Speaker="SuperEvo88" time="8:52:56 PM">
  <Utt Id="Utt273" Info-level="" Conventional="" Time="8:52:56 PM"
    Influence-on-listener="" Influence-on-speaker="" Agreement="Accept"
    Answer="">my iq right now is like in the negatives</Utt>
</Turn>
<Turn Id="T274" Speaker="AvrilLR" time="8:53:01 PM">
  <Utt Id="Utt274" Info-level="" Conventional="" Time="8:53:01 PM"
    Influence-on-listener="" Influence-on-speaker="" Agreement="None"
    Answer="">I had 10.39 instead of 10.392</Utt>
</Turn>
<Turn Id="T275" Speaker="AvrilLR" time="8:53:11 PM">
  <Utt Id="Utt275" Info-level="" Conventional="" Time="8:53:11 PM"
    Influence-on-listener="" Influence-on-speaker="" Agreement="Reject
    Part" Answer="">My IQ is 206...not joking...but that's off
    topic</Utt>
</Turn>
<Turn Id="T276" Speaker="AvrilLR" time="8:53:17 PM">
  <Utt Id="Utt276" Info-level="" Conventional="" Time="8:53:17 PM"
    Influence-on-listener="" Influence-on-speaker="" Agreement="None"
    Answer="">so now we add the two areas</Utt>
```



## 5. EVALUĂRI ȘI CONCLUZII

În urma testelor efectuate pe 3 fișiere exemplu (aproximativ 1200 replici), pentru prima abordare, bazată pe euristici, s-au obținut următoarele rezultate:

### *Statement*

- precizie 92%
- acoperire 79%

### *Info\_request*

- precizie 92%
- acoperire 92%

### *Action\_directive*

- precizie 67%
- acoperire 69%

Cea de-a doua abordare, bazată pe modele Markov, a recunoscut 80 % din actele de vorbire pe unul din fișierele cu dialoguri chat.

O primă îmbunătățire care poate fi adusă algoritmului euristic este recunoașterea părții de vorbire folosind unul din adnotatoarele existente gratuit pe web. Recunoașterea numelor utilizatorilor care apar în interiorul replicilor este o a doua problemă care va fi abordată în viitorul imediat, mai ales că majoritatea adresărilor pe un chat se fac cu prescurtări sau părți din numele vorbitorilor (nu cu numele complet).

Abordarea a doua este în curs de finalizare, în viitor extinzându-se numărul de acte de vorbire considerate.

## REFERINȚE

1. Allen, J., Core, M., Draft of DAMSL: Dialog Act Markup in Several Layers, <ftp://ftp.cs.rochester.edu/pub/packages/dialog-annotation/manual.ps.gz>
2. Austin, J.L., Cum să faci lucruri cu vorbe, Editura Paralela 45, 2003.
3. Boboc, Al., Limbaj și ontologie, Ed. Didactică și Pedagogică, 1997
4. Juravsky, D., Martin, J., Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Prentice Hall: San Francisco 2000
5. Manning, C., Schütze, H., Foundations of Statistical Natural Language Processing, MIT Press: Cambridge (Mass.) 1999
6. WordNet, <http://www.cogsci.princeton.edu/~wn>