Ana GUȚU, Universitatea Liberă Internațională, Chişinău, Republica Moldova

METODE MATEMATICE ÎN IERARHIZAREA TERMENILOR ŞTIINȚIFICI: SISTEMUL LIMBII ŞI THESAURUSUL

Înainte de a aborda subiectul construirii rețelei tip thesaurus [1] vom explicita noțiunile fundamentale de ***câmp terminologic*** şi ***sistem terminologic***. Pentru definirea lor vom opera cu noțiunile ***domeniu referențial*** (*DR*), ***bază de cunoştințe*** *(BC)* şi ***bază de date*** *(BD*). Vom menționa că DR este definit drept o reflectare denotativă în conştiința omului sau a colectivului a unui segment din realitatea obiectivă [2].

Dacă e să vorbim despre ingineria lingvistică, aici DR apare, în primul rând, ca o parte a lumii înconjurătoare, ce constituie sfera orientării euristice a intelectului artificial (IA) şi a procesării automate a textului (PAT). Descrierea sa este efectuată prin mijloacele limbii naturale (LN), care permite denumirea obiectelor, reflectarea particularităților acestora şi a raporturilor între ele.

Pentru sistemele automatizate sunt acceptabile doar descrierile în baza definițiilor DR. Astfel, limba formală trebuie co-raportată la limba naturală, ținând cont de natura deschisă, asociativ-dinamică, infinit-vagă a limbii naturale şi barierele ontologice ce o separă de limba fixă, rigidă, închisă a calculatorului [3].

Este destul de complicat de a defini Domeniul Referențial, în vederea diversității particularităților obiectelor lumii înconjurătoare, cât şi a inaccesibilității receptării complexe a întregului DR de către atenția directă a omului. În acest caz este necesară utilizarea metodei modelării. Modelarea Domeniului Referențial reprezintă o structurare a sistemului de noțiuni – reflectări ale obiectelor lumii înconjurătoare împreună cu particularitățile şi raporturilor lor. Vom aminti că noțiunea este o formă a gândirii, ce reflectă obiectele şi fenomenele lumii înconjurătoare împreună cu particularitățile şi raporturile lor esențiale. Noțiunea este o unitate a cunoaşterii. Sistema structurată a noțiunilor din perspectiva computerului constituie ***baza de date*** (BD). BD determină semantica intensională a modelelor (drept exemple de modele în prezentul articol servesc thesauri şi arborescențele). BD conține descrierea esențelor abstracte: obiecte, raporturi, procese. Esența abstractă – este noțiunea despre reprezentarea generalizată a unei clase de obiecte, raționamentul despre particularitățile sau raporturile dintre obiectele abstracte.

Cât priveşte noțiunea de BD, pe larg folosită în sistemele de procesare automată a textului (PAT) în anii 1960-1980, ea presupunea fixarea mai ales a semanticii extensionale a domeniului problematic, starea unor obiecte concrete, a unor parametri concreți pentru determinarea anumitor momente şi intervale de timp concrete.

Modelele de prezentare a cunoştințelor trebuie să ia neapărat în considerație non-exhaustivitatea principială a datelor. Prin urmare, modelarea DR este şi modelarea cunoştințelor, care se transformă şi se transmit sub formă de sistem de termeni. Sistematicitatea cunoştințelor este exprimată prin sistematicitatea termenilor.

În rezultatul modelării DR prin mijloacele limbii formale vom obține analogul, ce reflectă particularitățile noționale esențiale ale acestui DR. Un astfel de model structurat sau designativ al DR, desfăşurat la nivelul noțional, va fi numit în continuare ***Spațiu Semantic*** (SS). În terminologie acestei noțiuni îi corespunde conceptul de câmp terminologic. Câmpul terminologic este subdivizat în sectoare şi micro-câmpuri până la nodurile terminale, adică semnificațiile unor termeni aparte şi raporturile ce unesc aceste semnificații. Numărul sectoarelor semantice, frontierele lor, desigur, sunt clar delimitate de însăşi mijloacele limbii [4].

Câmpul terminologic este o entitate sistemică de esență conținutistică. Pentru planul expresiei vom adopta noțiunea de ***sistem terminologic*** ca simbioză organizată a unităților terminologice, cât şi a mijloacelor formale sintactice, ce exprimă raporturile sintagmatice şi

BDD-A3399 © 2009 Editura Europlus
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-19 21:32:56 UTC)

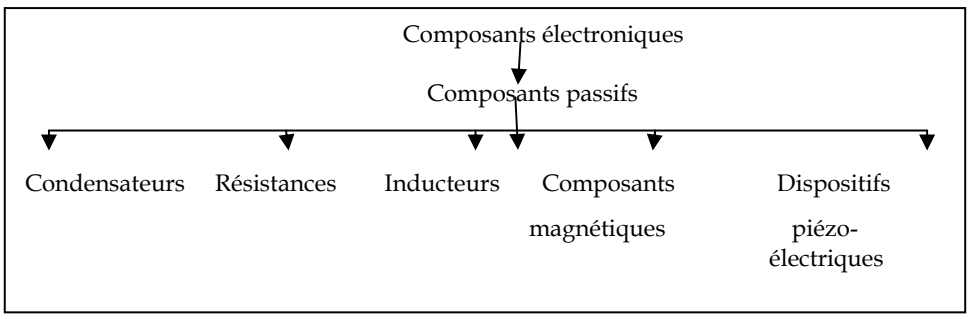
paradigmatice între unitățile terminologice. Combinarea câmpului terminologic (semnificantului) şi a sistemului terminologic (semnificatului) ne dă ca rezultat un construct semnic diadic, numit ***sistemul terminologic lexico-semantic (STLS)***. Aşa dar, prin STLS vom înțelege un semn complex, a cărui unitate este asigurată de unitatea dimensiunii conținutului, adică a DR respectiv.

Deoarece omogenitatea structurală a STLS este asigurată de planul conținutului, pentru construirea lui se purcede la structurarea câmpului terminologic, iar apoi sunt selectate unitățile terminologice, ce formează sistemul terminologic.

Fiecare DR poate fi divizat în noțiuni non-subdivizibile. Descrierea structurii DR presupune indicarea tipurilor principale ale raporturilor între sectoare, microcâmpuri şi nodurile câmpului terminologic, cât şi definirea locului precis al tuturor acestor obiecte în sistem.

Din toate metodele existente - semiotice, lingvo-matematice şi tradițional-lingvistice, utilizate pentru construirea designativă a câmpului terminologic, destul de eficace s-a dovedit a fi metodologia thesaurus-rețea, ce se bazează pe noțiunea de mulțime, algebra raporturilor şi teoria grafurilor existențiale. Utilizarea sa în terminologie cere o concentrare prealabilă a obiectelor lingvistice vagi [5] şi aducerea entităților lingvistice incerte la precizia elementelor mulțimilor discrete [6]. Vom aminti noțiunile de bază utilizate în metodologia thesaurus-rețea. Aici noțiunea de bază este cea a mulțimii (A), compuse din elemente (x): A = $\{x_1, x_2, ...x_i, x_j, ...x_{n-1}, x_n\}$, unde toate elementele (x) aparțin mulțimii A. Mulțimea ce nu conține nici un element, este considerată vidă: A = Ø. Produsul cartezian A × A este mulțime, ce constă din toate perechile de elemente organizate ale mulțimii A, adică A × A = $\{(x_1, x_2), ..., (x_1, x_i), ... (x_i, x_j), ..., (x_{n-1}, x_n)\}$, unde orice submulțime ρ, ce aparține mulțimii A × A, se numeşte raport binar.

Vom cita un exemplu de mulțime lingvistică în funcționarea metalimbajului „Materiale şi componente ale tehnicii de calcul". Mulțimea A conține ULT, ce țin de segmentul *passive components*. În desenul No1 este reprezentat un graf, ce descrie raporturile dintre elementele acestei mulțimi.



Desenul No1. Raporturile binare între elementele mulțimii A a câmpului terminologic „Materiale şi componente ale tehnicii de calcul"

Atunci avem:

A = {*composants électroniques, composants passifs, condensateurs, résistances, inducteurs, composants magnétiques, dispositifs piézo-électriques*}.

În mulțimea A să instituim raportul binar ρ_h " a fi în raport semantic „specie-gen"". Acest raport binar conține perechile:

BDD-A3399 © 2009 Editura Europlus
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-19 21:32:56 UTC)

ρ_h = {*(composants électroniques, composants passifs), (composants passifs, condensateurs), (composants électroniques, condensateurs), (composants passifs, résistances), (composants électroniques, résistances), (composants passifs, inducteurs), (composants électroniques, inducteurs), (composants passifs, composants magnétiques), (composants électroniques, composants magnétiques), (composants passifs, dispositifs piézo-électriques), (composants électroniques, dispositifs piézo-électriques)*}.

La construirea organizării ierarhice STLS sunt folosite astfel de raporturi binare între elemente, cărora le este caracteristică ***tranzitivitatea***. În cazul raporturilor de tranzitivitate elementele pot fi structurate doar într-o consecutivitate strict determinată, cu condiția lipsei reflexivității, ce ar permite elementelor să se auto-substituie, adică să se obțină perechile (x, x) şi (y, y) şi a simetriei, ce ar permite formarea perechilor identice de tipul (x, y) şi (y, x) [7]. La prezența simetriei şi reflexivității rețelele vor fi simple (fără ierarhie, adică fără raporturi de subordonare).

Rețelele ierarhice se reprezintă fie prin verigi (fiecare nivel al ierarhiei este materializat printr-un singur element), fie prin thesaurus. Anume rețelele ***thesaurus*** denotă un interes pentru soluționarea problemelor ingineriei lingvistice. Pe lângă aceasta, noi vom examina şi raporturile, utilizate în rețelele non-ierarhice (acestea din urmă permit construirea unui model thesaurus şi a oricărui DR şi mai eficace).

Aşa dar, nucleul modelului thesaurus al fiecărui DR este construit sub formă de rețea ierarhică, care are aspectul unei mulțimi organizate de elemente lexicale, ce sunt corelaționate prin raporturi binare subordinate. Caracterul subordonat al raporturilor este transmis prin codul nodului, atribuit fiecărui cuvânt-termen sau grup de cuvinte terminologic. Să examinăm trăsăturile principale ale rețelei thesaurus în baza exemplului metalimbajului „Materiale şi componente ale tehnicii de calcul" (desenul No2):

1. La ***rădăcina arborescenței*** se află termenul *composants électroniques* – componente electronice (este noțiunea cea mai abstractă în raport cu cele derivate, noțiunea generică).

2. Celelalte noduri se conțin în mulțimi binare ce nu se intersectează, fiecare din ele fiind o sub-arborescență sau arborescență a rădăcinei respective) – *composants passifs* – componente pasive (condensateurs – condensatori, *résistances* – rezistori, *inducteurs* – inductori, *composants magnétiques* – componente magnetice, *dispositifs piézo-électriques* – dispozitive piezo-electrice, *composants actifs* – componente active, (*tubes électroniques* – lămpi electrice, *dispositifs à semiconducteurs* – dispozitive semiconductoare, *circuits intégrés* - scheme integrale), *dispositifs opto-électroniques* – dispozitive optoelectronice, (*diodes émettrices et lasers à semiconducteurs* – diode emițătoare şi lazere cu semiconductori, *photodétecteurs* – fotodetectori, *dispositifs photovoltaïques* – dispozitive fotovoltaice, *piles solaires* – baterii solare).

3. Nodul, de la care nu derivă arborescențe, se numeşte ***nod terminal (punct final).***

4. Arborescenței îi pot fi aplicate câteva raporturi binare, bazate pe tranzitivitate, antireflexivitate şi asimetrie.

Aceste caracteristici sunt fundamentale pentru stabilirea în thesaurus a raporturilor „**specie-gen"** şi **„parte-întreg".**

În thesaurusul nostru drept exemplu de raporturi tranzitive, antireflexive şi asimetrice serveşte următoarea dependență:

composants électroniques → *composants passifs* → *condensateurs*.

Într-adevăr, acest raport este tranzitiv, deoarece perechile (*composants électroniques, composants passifs)* şi (*composants passifs, condensateurs*) presupun existența perechii (*composants électroniques, condensateurs*). Adică toate cele trei perechi denotă raporturi de tip „gen-specie".

Tripletul de termeni analizați este *antireflexiv*, deoarece în thesaurus lipsesc perechile ce reprezintă în sine reluarea termenului, de exemplu (*composants électroniques, composants électroniques)*.

Asimetria constă în aceea că, de exemplu, prezența perechii *composants électroniques, composants passifs*, încadrată în raportul gen-specie, nu permite un raport reversibil: *composants passifs (specie), composants électroniques (gen)*. Cu alte cuvinte, mişcarea în

BDD-A3399 © 2009 Editura Europlus
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-19 21:32:56 UTC)

thesaurus este posibilă doar de sus în jos, trecând de la trunchiul arborescenței spre segmente derivate ce nu se intersectează., micro-câmpuri, domenii şi sub-domenii, ce concretizează pas cu pas fiecare noțiune a DR.

Arborescențele sunt pasibile şi altor raporturi. De exemplu, raportul de antonimie (*composants passifs - composants actifs* / componente pasive – componente pasive) nu este un raport tranzitiv (această pereche nu presupune prezența unui termen intermediar), dar este un raport simetric (el prevede reversibilitatea sa) şi totodată este un raport antireflexiv (acest raport principial nu poate conține perechi de elemente lingvice identice). Raportul de asociație după asemănare (*céramique - verre - films oxydés* - ceramică, sticlă, pelicule oxidate) sunt caracterizate prin tranzitivitate şi antireflexivitate. Rețelele construite în baza raporturilor non-ierarhce, nu pot fi numite arborescente. Le vom întitula pur şi simplu *rețele*. Din punct de vedere schematic ele pot fi reprezentate sub formă de linii, la capătul cărora se situează noțiuni independente una față de alta sau sub formă de poliedru închis.

[2]Composants électroniques - componente electronice		
↓ → [2.1] Composants passifs Componente passive	↓ → [2.2] Composants actifs Componente active	↓ → [2.3] Dispositifs optoélectroniques Dispozitive optoelectronice
→[2.1.1] Condensateurs Condensatori	→[2.2.1] Tubes électroniques Lămpi electronice	→[2.3.1] Diodes émettrices et lasers à semi-conducteurs Diode emițătoare cu laser şi semiconductoare
→[2.1.3] Résistances Rezistoare	→ [2.2.2] Dispositifs à semi-conducteurs Dispozitive semiconductoare	→[2.3.2] Photodétecteurs Fotodetectori
→ [2.1.4] Inducteurs Inductori	→ [2.2.3] Circuits intégrés Scheme integrale	→ [2.3.3] Dispositifs photovoltaïques, piles solaires Dispozitive fotovoltaice, baterii solare
→ [2.1.5] Composants magnétiques Componente magnetice		
→ [2.1.6] Dispositifs piézo-électriques Dispozitive piezoelectrice		

Desenul No5. Fragment din thesaurusul „Materiale şi componente ale tehnicii electronice".

BDD-A3399 © 2009 Editura Europlus
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-19 21:32:56 UTC)

Caracteristicile raporturilor binare – ***tranzitivitatea, antireflexivitatea şi asimetria*** stau la baza construirii reţelelor thesaurus nu doar pentru metalimbajele strict organizate. O reţea thesaurus poate fi aplicată şi la modelarea metalimbajelor mai puţin structurate precum cele ale DR „Moda" şi „Cosmetica". Fragmentul din thesaurusul DR „Moda"demonstrează în ce mod se concretizează raporturile între noţiuni. De exemplu, *veston* –veston - este un termen mai generic decât *smoking* – redngotă, dar nu este legat de el printr-un raport direct, deoarece ţine de ramificaţia *veste de costume* – veston clasic, de la costum, şi nu de ramificaţia *veste habillée* – veston de gală. În mod analogic sunt construite reţelele thesaurus pentru metalimbajele „Economie" , „Regnul animal" ş.a.

Drept concluzie putem afirma că, ***termologia (STLS) este o entitate semnică ce include câmpul terminologic (semnificatul) şi sistemul terminologic (semnifcantul). Mai sus am menţionat deja că drept model al vocabularului unui DR poate apărea nu doar reţeaua arborescentă, construită în exclusivitate în baza raporturilor ierarhice, ci şi reţeaua, ce cuprinde pe lângă raporturile subordonate şi cele non-ierarhice.***

Vom menţiona: câmpurile terminologice reprezentate sub formă de reţele thesaurus, sunt de obicei universale cu referire la diferite limbi. În acest caz pentru desemnarea nodurilor de reţea este judicioasă folosirea unui limbaj codat (codul cifrelor, literelor) sau a unei limbi naturale. Iată de ce în lucrările ştiinţifice la temă sunt folosite pentru notarea acestor noduri termeni francezi, mai rar englezi. Recurgerea la limba engleză se explică prin caracterul diferenţiat al terminologiei sale în raport cu limbile romanice.

NOTE

[1] Reţeaua thesaurus este instrumentul lexicografic ce face faţă cel mai bine sarcinii de sistematizare a aparatului terminlogic dintr-un anumit dmeniu al cunoaşterii. Thesauri sunt cnstruite în baza analizei lexico-statistice a mesajelor textuale şi reprezintă culegeri de terminlogii standardizate şi reglementate. Cuvântul ***thesaurus*** (din greacă θησαυρός – comoară, tezaur) este folosit în sensul larg pentru a desemna cartea ce conţine o colecţie sufficient de exhaustivă de noţiuni, denumiri din diverse domenii ale cunoaşterii. Cât priveşte teoria şi practica căutării informaţionale, cuvântul thesaurus este utilizat pentru denumirea dicţonarelor de noţiuni şi clasificarea raporturilor între ele. În comparaţie cu dicţionarul obişnuit, thesaurusul este mai flexibil şi ma adaptat la descrierea semantică a limbi, graţie diversităţii şi profunzimii raporturilor stabilite între noţiunle terminologice.

[2] Piotrowski R., Popeskul A., Chažinskaja M., Rachubo N. Automatische Wortanalyse. Bochum: Brockmeyer, 1985, p.45.

[3] Piotrowski R., Zaitseva N. Behavior-based artificial intelligence and new machine-translation conceptions // Human and Computer. Verbal communication and interaction via computer. 9th Conference of GRLA-RWCAL. Bacău-Tescani, 26-29 april 2001. Constanţa: Editura Europolis, 2001, p.205-211.

[4] Marcus S., Nicolau Ed., Stati S. Introducere în lingvistica matematică. Bucureşti: Editura ştiinţifică, 1966, p. 17; Anzaldi C., Bordoni L., Sano A. Construction of a Terminological Interdisciplinary Thesaurus // TKE'96. Terminology and Knowledge Engineering. Proceedings: Fourth International Congress on Terminology and Knowledge Engineering, 26-28 August 1996, Vienna, Austria, Frankfurt a. M.P. 273-278; Toft B. Conceptual Relations in Terminology and Knowledge Engineering // International Conference on Terminology Science and Terminology Planning. Riga 17-19 August 1992 and International IITF-Workshop Theoretical Issues of Terminology Science. Riga 19-21 August 1992, Vienna 1994. P. 259-282.

[5]Hubey M.H. Mathematical Foundation of Linguistics. München: LINCOM Handbooks in Linguistics, 2000. (Secţiunile: 20, 24, 40); Tieping WU. Fuzzy linguistics and terminology // TermNet News. Journal of the International Network for terminology, published in Austria, Wien. 1989. N 25. P. 25-27.

[6] Malvache N., Vidal P. Application des systèmes flous à la modélisation des phénomènes de prise de décision et d'appréhension des informations visuelles chez l'homme. P.: A.T.P. - C.N.R.S. 1 KO5, 1974.

[7] Toate aceste raporturi binare pot fi reprezentate in baza exemplului submulţimilor unor mulţim non-ierarhice $A \times A = \{(x, y, z) (x, y, z)\}$, adică

BDD-A3399 © 2009 Editura Europlus
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-19 21:32:56 UTC)

A × A = {(x, x) (x, y) (x, z) (y, y) (y, x) (y, z) (z, z) (z, x) (z, y)}. Atunci submulţimea sa Br, Bs, Bt demonstrează raportul

Reflexivităţii Br = {(x, x) (y, y) (z, z)},

Simetriei Bs = {(x, y) (y, x) (x, z) (z, x) (y, z) (z, y)},

Tranzitivităţii Bt = {(x, y) (y, z) (x, z)}.

SURSE BIBLIOGRAFICE

Anzaldi C., Bordoni L., Sano A. (1996). *Construction of a Terminological Interdisciplinary Thesaurus* // TKE'96. Terminology and Knowledge Engineering. Proceedings: Fourth International Congress on Terminology and Knowledge Engineering, 26-28 August 1996, Vienna, Austria, Frankfurt a. M.P. 273-278.

Hubey M.H. (2000). *Mathematical Foundation of Linguistics*. München: LINCOM Handbooks in Linguistics,. (Secţiunile: 20, 24, 40).

Malvache N., Vidal P. (1974). *Application des systèmes flous à la modélisation des phénomènes de prise de décision et d'apprehension des informations visuelles chez l'homme*. P.: A.T.P. - C.N.R.S. 1 KO5

Marcus S., Nicolau Ed., Stati S. (1966). *Introducere în lingvistica matematică*. Bucureşti: Editura Ştiinţifică, p. 17.

Piotrowski R., Popeskul A., Chažinskaja M., Rachubo N. (1985). *Automatische Wortanalyse*. Bochum: Brockmeyer, p.45.

Piotrowski R., Zaitseva N. (2001). *Behavior-based artificial intelligence and new machine-translation conceptions* // Human and Computer. Verbal communication and interaction via computer. 9th Conference of GRLA-RWCAL. Bacău-Tescani, 26-29 april 2001. Constanţa: Editura Europolis, p.205-211.

Tieping WU. (1989). *Fuzzy linguistics and terminology* // TermNet News. Journal of the International Network for terminology, published in Austria, Wienna. 1989. N 25. P. 25-27.

Toft B. (1992). *Conceptual Relations in Terminology and Knowledge Engineering* // International Conference onTerminology Science and Terminology Planning. Riga 17-19 August 1992 and International IITF-Workshop Theoretical Issues of Terminology Science. Riga 19-21 August 1992, Vienna 1994. P. 259-282.

ABSTRACT

In this article the author describes the principles of the structure of technical terminology in French language in basis of the thesaurus branches. The thesis is that the ***terminological system*** *and the* ***terminological field*** *are two main parts of the each referential domain: engineering, electronics, computers, telephony etc. The structure of the thesaurus branches is a signs entity, that may be transitive, subordinate, symmetrical (antonymic), non-symmetrical, anti-reflexive. The branches that are non-hierarchical can not be named thesaurus branches. The thesaurus branches are characteristic to other domains with a little degree of organization like Mode, Economics, Fauna. The thesaurus branches are an universal phenomenon developed in different languages, that means we can speak about the existence of the inter****-lingual conceptual network****.*

BDD-A3399 © 2009 Editura Europlus

Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-19 21:32:56 UTC)