No: 02 2025
https://doi.org/10.64400/ndueej.039
ISSN 2224-5529 2224-0829 2223-5124 2222-940X

NAXÇIVAN DÖVLƏT UNİVERSİTETİNİN ELMİ ƏSƏRLƏR JURNALI

FRANSIZ DİLÇİLİYİNDƏ ONOMASTİK VAHİDLƏRİN AVTOMATİK İDENTİFİKASİYA XÜSUSİYYƏTLƏRİ

Vəfa Seyid

https://orcid.org/0000-0003-0725-0232
Azərbaycan Dillər Universiteti, R. Behbudov küç., 134, Bakı, Azərbaycan
**Yazışılan müəllif:* vafaseidova@mail.ru; *Tel.: +994 50* 6804080

XÜLASƏ

Məqalə fransız dilçiliyində xüsusi isimlərin avtomatik tanınması xüsusiyyətlərini araşdırır, xüsusi isimləri ümumi isimlərdən fərqləndirən xüsusiyyətləri tədqiq edir və xüsusi ismin ayrıca bir fərd və ya obyekt kimi mövcudluğu ilə bağlı müxtəlif dilçi alimlərin baxışlarını təhlil edir. Qeyd olunur ki, mətn daxilində xüsusi isimlərin tanınması bəzən dilçilər üçün çətinliklər yaradır. Avtomatik dil işlənməsi baxımından, xüsusi isimlərin tanınması üzrə aparılan işlər kompüter mütəxəssislərini kompüter işi üçün daha sadə tiplər və praktik həllər təklif etməyə vadar edir, eyni zamanda xüsusi isimlərin reallığını yetərincə nəzərə almağa çalışır. Xüsusi adların çıxarılması, tanınması və kateqoriyalaşdırılması üçün üç tip obyektin fərqləndirilməsi təklif olunur: **ENAMEX, TIMEX** və **NUMEX**. Xüsusi isimləri ayırd etmək üçün üç əsas sistem tipi təqdim olunur: **qaydalara əsaslanan sistemlər, öyrənməyə əsaslanan sistemlər** və **hibrid sistemlər.** Məqalədə göstərilir ki, xüsusi isimlərin aşkarlanması və təsnifatı üçün ən etibarlı ipucu onların **sol və ya sağ konteksti** və **daxili quruluşudur**. Xüsusi isimlər cümlə quruluşunda epitet, təyin, mübtəda, tamamlıq və ya təsrifləyici kimi də çıxış edə bilər. Tədqiqat göstərir ki, müxtəlif tipli xüsusi isimlər identifikasiya baxımından bərabər deyil, çünki onların kontekstual görünüşü və qəzet məqalələrində rastgəlmə tezliyi əhəmiyyətli dərəcədə fərqlənir. Buna görə də, xüsusi isimlərin identifikasiyası üçün istifadə olunan alətlər bu fərqlərə uyğunlaşdırılmalıdır.

Açar sözlər: *avtomatik identifikasiya, hibrid sistemlər, delimitasiya, variasiya, daxili və xarici sübut, kontekst, kateqoriya.*

GİRİŞ

Xüsusi ismin ümumi isimdən fərqli cəhətlərinin tədqiq edilməsi, onun ayrıca bir fərd və ya obyekt kimi mövcud olması haqqında dilçi alimlər müxtəlif mülahizələr irəli sürmüşlər. Tədqiqatlar zamanı biz hətta xüsusi isimlərin bəzi dilçilər tərəfindən əsl xüsusi isimlər (*Jean-Jaques Rousseau, Franche-Compté*) və qarışıq (xüsusi isim və ümumi ismin, hətta sifətin birləşməsi – *île de France, l'université de Sorbonne, la Nouvelle Calédonie*) xüsusi isimlər kimi növlərə ayrılmasının da şahidi olduq.

Lakin xüsusi isimlərin mətn daxilidə tanınması bəzən dilçilər üçün çətinliklər yaradır. Avtomatik dil emalı cəhətdən, xüsusi adların tanınması üzərində iş kompüter alimlərini daha sadə tipologiyalar və kompüter işi üçün praktiki istifadəsi olan, lakin xüsusi adların reallığını kifayət qədər nəzərə alan təklif verməyə vadar etdi.

Oxucu xüsusi adı tanımırsa, ümumi diskurs onlara bu adı tanımağa, həm də onun hansı növ olduğunu anlamağa kömək edəcək: yer, şəxs və ya başqa ad. Oxucu üçün xüsusi adın tanınmasının iki səviyyəsi var ki, bunlar bir-birini istisna etmir: ya xüsusi ad məlum olduğu üçün tanınır və o, ümumi kainat biliyinə aiddir (məsələn: Luara, Paris, Sartr), ya da xüsusi adın növünü yaradan və ya qeyri-müəyyənlik halında onu təyin edən predikatların yazılışı (böyük hərfin olması) və semantikasıdır.

Əslində, 1987-ci ildə MUC6 (Message Understanding Conference, <http: //www.muc.saic.org>) tədqiqat proqramının yaradılmasınadək kompüter elmində xüsusi isimlərə çox kiçik tədqiqatlar yönəldilmişdi. Bu proqramın məqsədi mətnin avtomatik başa düşülməsində tədqiqatı təşviq etməkdir. MUC, iştirakçı sistemlərin qiymətləndirildiyi müsabiqə formasını alır.

MUC tərəfindən təklif olunan əsas vəzifə suallara cavab vermək üçün mətnlərdə olan məlumatların çıxarılmasıdır: kim?, nə vaxt?, harada?, nə?, necə?

Adlandırılmış obyektlərin çıxarılması, tanınması və kateqoriyalara ayrılması üçün üç növ obyekti ayırmaq təklif olunur: ENAMEX, TIMEX və NUMEX. TIMEX vaxt və tarixi göstərən ifadələri, NUMEX rəqəmləri və faizləri, həmçinin pul kəmiyyətlərini bir araya gətirir, ENAMEX xüsusi adlar və akronimlərdən ibarətdir. Xüsusi isimləri çıxarmaq üçün üç əsas sistem növü var:

Qaydalara əsaslanan sistemlər: Sistemlərin əksəriyyəti bu yanaşmadan istifadə edir. Tipik qaydalara əsaslanan sistemlər xüsusi adları (böyük hərf, konkret sözün olması və s.), eləcə də artıq məlum olan xüsusi adların lüğətlərini müəyyən etmək üçün linqvistik təsvirlərdən və ipuclarından istifadə edir. Qaydalar əl ilə yazılır. Bu sistemlər çox yaxşı nəticələr verir, lakin əhəmiyyətli dərəcədə insan sərmayəsi tələb edir. Bu tip strategiya ciddi redaksiya meyarlarına cavab verməyən mətnlər (məsələn, e-poçtlar) üçün ideal deyil.

Öyrənmə sistemləri: Təlim korpusunda öyrənməklə avtomatik olaraq xüsusi adlar haqqında bilikləri artırmaq olur. Bu sistemlər bütün növ mətnlərə tez uyğunlaşdırıla bilər, lakin qaydalara əsaslanan sistemlərdən dəqiqliyi daha az olur. Öyrənmə sistemləri təsvir işini minimuma endirir, lakin daha aşağı nəticələr verir.

Hibrid sistemlər: Onlar əl ilə yazılmış qaydalardan istifadə edir, eyni zamanda öyrənmə alqoritmlərindən istifadə edərək təlim məlumatlarından götürülmüş sintaktik məlumat və diskurs məlumatlarından istifadə edərək öz qaydalarının bir hissəsini yaradırlar. Avtomatik xüsusi ad çıxarma sisteminin nəticələrini qiymətləndirmək üçün iki ölçüdən istifadə olunur. *Xatırlatma* sistemin ideal cavabların sayına nisbətdə malik olduğu düzgün cavabların miqdarını ölçür. *Dəqiqlik* sistemin verdiyi bütün cavablar (düzgün və səhv) arasında düzgün cavabların miqdarıdır.

Müxtəlif MUC sistemləri tərəfindən göstərilən nəticələr çox yaxşıdır, lakin yadda saxlamaq lazımdır ki, onlar kifayət qədər məhdud domenlə məhdudlaşan çox homojen mətnlərlə məşğul olurlar (məsələn: *des dépêches AFP* – AFP xəbərləri). Le Monde qəzetlərinin korpusunda sınaqdan keçirilmiş, Prolex layihəsi üçün yaradılmış ExtracNP sistemi fransız dilində bu an üçün ən yaxşı nəticələri əldə etmişdir: le Monde qəzetindəki xüsusi isimlərin 93,2%-i 94,4% dəqiqliklə təyin edilir. Bu, Intex sisteminin çevirici formalizmindən istifadə edən qayda-əsaslı sistemdir.

Qayda əsaslı sistemlə xüsusi isimləri necə tanımaq olar?

Xüsusi adların tanınması və yazılması kəsişən iki problemdir. Doğrudan da, xüsusi adı çıxarmaq üçün onu müəyyən etməyə imkan verən ipuclarından istifadə etmək, həm də onu kateqoriyalara ayırmaq lazımdır. Yalnız sintaksisdən istifadə edən çıxarma sistemi xüsusi isimlə ümumi isim arasında fərq qoya bilməz və xüsusi ada kateqoriya təyin edə bilməyəcək. Xüsusi adların sistematik cəhəti və strukturu var ki, onu sintaktikdən daha çox leksik olan məlumatlardan istifadə etməklə təsvir etmək olar. Xüsusi isimləri çıxarmaq üçün ilk sadə ipucu böyük hərfin olmasıdır: bu, kifayət deyil, çünki xüsusi isim bir neçə sözdən ibarət ola bilər və onların bəziləri böyük hərflə yazılmır. Bundan əlavə, cümlənin ilk sözündə yazılan böyük hərf birmənalı deyil: bu xüsusi isimdir, yoxsa sadəcə olaraq, cümlənin əvvəlində olduğu üçün böyük hərflə yazılmış bayağı sözdür?

Əslində, xüsusi isimləri aşkar etmək və təsnif etmək üçün ən etibarlı ipucu onların görünüşünün sağ və ya sol kontekstləri və daxili quruluşudur.

Daxili və xarici sübutlar

Xüsusi adın tanınması üçün əksər kompüter vasitələri bu sübutları belə adlandırmadan istifadə edir.

Daxili sübutlar xüsusi adın öz daxilində tapılır. Bunlar onu dəqiqliklə müəyyən etməyə və bəlkə də onu yazmağa imkan verən sözlərdir. Daxili sübut bir və ya bir neçə sözdən və ya xüsusi adın bir hissəsi olduğu bilinən abreviaturadan ibarət ola bilər (məs: *NATO=Organisation du traité de l'Atlantique nord* – Şimali Atlantika Müqaviləsi Təşkilatı, *le Pont Neuf* – Yeni Körpü, *Nouveau Journal d`Onomastique* – Yeni Onomastika jurnalı). Belə sözlərə xüsusi isimlərin əvvəlində və ya sonunda rast gəlinir (xüsusilə təşkilat adlarında). İnsan adı da daxili sübut kimi istifadə edilə bilər (Məs: *Victor Hugo* Viktor Hüqo).

Xarici sübut xüsusi isimlərin cümlədə işlənmə kontekstdir. Xüsusi isimlər müəyyən bir növ fərdlərə müraciət etmək üsuludur. Müəllif nitqində, xüsusən də jurnalistikada istinad etdiyi insanlar, yerlər, təşkilatlar haqqında oxuculara əlavə məlumat verir: bu məlumatlar avtomatik tanıma prosesində xüsusi ismin növünü təyin etməyə kömək edə bilər. Xarici sübut həm də kontekstin cümlədəki xüsusi ismin sağında və ya solunda olub-olmamasından asılı olaraq sağ kontekst və ya sol kontekst adlandırılır: *la ville de Lion* Lyon şəhəri, *le groupe Gazelli* Qazelli qrupu, *brand Levis* Levis brendi, *la police française* fransız polisi və s.

Xüsusi adlar cümlədə konstruksiyaların tərkibində epitet, atribut, subyekt, obyekt, təyin funksiyasında da ola bilərlər. Xüsusi isimlər mürəkkəb konstruksiyaların kontekstində sadəcə olaraq sifət ola bilər (məsələn: *anglaise Margaret Tatcher* ingilis Margarita Tetçer) və ya daha mürəkkəb formada ola bilər (məsələn: *Le chef du gouvernement français Nicolas Sarkozy* Fransa hökumətinin başçısı Nikolay Sarkozi).

Xüsusi isimlərin variasiyası

Xüsusi adları tanımaq üçün onların Daille və Morin (2000) aşağıdakı kimi sadaladığı variasiyalarını nəzərə almaq lazımdır: qrafik variantlar (məs.: Kommunist Partiyası və ya kommunist Partiyası), akronimlər və ya abreviaturalar kimi variantlar, müəyyən koordinasiyalar (məsələn: *le grand bureau et le petit bureau → le grand et le petit bureaux* böyük idarə və kiçik idarə→ böyük və kiçik idarə), ellipslər (məsələn: *école privée sup. → privée sup. → privée* özəl məktəb → özəl ali → özəl) [115, s.601-621].

Xüsusi isimlərin delimitasiyası

Jacquemin və Bush (2000) adlandırılmış obyektlərin zəif təsviri ilə bağlı səhvləri aşağıdakı kimi müəyyən etmişlər:

− Həddindən artıq tanınma: tanınan ardıcıllıq adı çəkilən obyekti ehtiva edir, lakin çox uzundur.

− Yetərsiz tanınma, tanınan obyektin orijinal obyektdə olması faktını təsvir edir. Məsələn, *L'ancien président George Pompidou a visité les pays africains* Keçmiş prezident Jorj Pompidu Afrika ölkələrinə səfər etdi cümləsində, əgər biz yalnız Jorj Pompiduya diqqət yetirsək, obyekt az tanına bilər, çünki biz Jorj Pompidunu axtarmalı olarıq.

Həddindən artıq tanınma və zəif tanınma xüsusi isimlərin sağ hissəsində xüsusilə özünü büruzə verir. Biz xüsusi ismin başlanğıcını olduqca asanlıqla tapırıq (böyük hərfin olması), lakin ondan sonrakı sözlər baş hərflə yazılmaya bilər; deməli, düzgün həddi tapmaq çətindir (məsələn: *La Fédération nationale de la Mutualité française* Fransız Qarşılıqlılıq Milli Federasiyası) [150, s.181-189].

Volınski və başqaları (1995) bu problemi qrammatik markerlərdən (sözönülər, bağlayıcılar, vergüllər, nöqtələr) istifadə edərək xüsusi isimləri seqmentlərə ayıraraq qismən həll edir, lakin bu seqmentasiya qeyri-kafi olur, çünki xüsusi isimlər bağlayıcılar və ya ön sözönülərdən ibarət ola bilər [234, s.23-30].

Xüsusi ismin sağ hissədən genişlənməsi xüsusi ismin növündən asılı olan və mütləq isim və ya sifətlə bitən sifətlər, isimlər, ön sözlər, təyinedicilər ehtiva edə bilər. Sağa uzadılma imkanları nəzərdən keçirilən xüsusi ismin növündən asılıdır. Bununla belə, sifətin bu və ya digər növ xüsusi isimdən sonra işlədilməsi mümkün olsa belə, bu, problem yaradır: *L'Amérique centrale* Mərkəzi Amerika bir varlığı ifadə edir, lakin *Amérique riche* zəngin Amerika sözündə yalnız *Amérique* Amerika xüsusi isimdir və *zəngin* sözü onun bir hissəsi deyil.

Digər struktur qeyri-müəyyənliyi cümlənin əvvəlində və nöqtədən sonra baş hərfin olması ilə bağlıdır: Bu böyük hərf cümlənin başlanğıcını göstərirmi? Xüsusi isimdirmi? Hər ikisidirmi ya yoxsa heç biri? Buna görə də Silberzteyn və başqa tədqiqatçılar bildirirdilər ki, böyük hərfi və nöqtəni ayırd etmək və beləliklə, cümlənin əvvəlini və sonunu bilmək üçün mətnləri cümlələrə bölmək lazımdır [220, s.269-276].

Böyük hərflər və ya rəqəmlər olduqda nöqtə birmənalı deyil. Cümlə başlanğıclarından başqa, həm böyük hərflər, həm də nöqtələrdən ibarət nümunələr dörd növə ayrılır:

– Şəxs adları, onlardan əvvəl qısaldılmış fəxri adlar, sivil adlar (məsələn: ***M.*** *Dupont* cənab Düpon, ***Mme*** *Durand* xanım Düran) və ya onlardan əvvəl qısaldılmış ad (məsələn*:* ***J.*** *Dupont* J.

Düpon) olduqda.

– Akronimlər (məsələn: **S.N.C.F** (*Société nationale des chemins de fer français* – Fransa milli dəmiryol şirkəti).

– Böyük hərflə bitən mürəkkəb sözlər və simvollar (məsələn: ex.: Ce bijou coûte 30 **F.** Il a été acheté chez un bijoutier. Bu bəzək əşyasının qiyməti 30 frankdır. O, zərgərdən alınıb).

– Müxtəlif abreviaturalar (məsələn: éd. Clé international – Beynəlxalq Kle redaksiyası).

– Bir böyük hərfdən ibarət olan simvollar: ölçü vahidlərinin abreviaturaları (məsələn: V = volt, W = vatt və s.) və valyuta simvolları (məsələn: F = Frank və s.) cümlə daxilində heç bir problem yaratmır, çünki onlardan sonra nöqtə qoyulmur (məsələn: Luvr muzeyindən 50.000 F dəyərində portret oğurlanıb). Lakin onlar cümlənin sonunda olduqda, simvollardan sonra ardıcıllığı qeyri-müəyyən edən bir nöqtə gəlir: ardınca nöqtə olan simvol adın başlanğıcı kimi və onun ardınca gələrək böyük hərflərlə yazılan söz, soyad kimi hesab edilə bilər (məsələn: *C'est un tableau de Monet de 50 000 F. Volé au Musée de Louvre, il ne sera sans doute jamais retrouvé* Bu, 50.000 F dəyərində olan rəsmdir. Luvr Muzeyindən oğurlanmışdır, yəqin ki, heç vaxt tapılmayacaq).

Dil ipuclarından istifadə edərək müəyyən edə və təsnif edə biləcəyimiz xüsusi isimlərin miqdarını daha yaxşı başa düşmək üçün bir araşdırma apardıq. Le Monde qəzetinin bir nömrəsində apardığımız araşdırmada biz xarici və daxili dəlilləri müşayiət etdikləri xüsusi isimlərin növünə görə saydıq. (Ce travail a été réalisé sur un journal *Le Monde* complet, daté du 12 janvier 1999 – Bu iş Mond qəzetinin1999-cu il 12 yanvar tarixli nəşri üzərində həyata keçirilmişdir).

Aşağıdakı kateqoriyalardan dilçiliyə və xüsusi isimlərin avtomatik tanınmasına aid cəmi 3755 xüsusi isim tapdıq: insanlar (27%), yerlər (35,2%) və qəbilələr (8,4%), təşkilatlar (27%); obyektlər və markalar, hadisələr və fəlakətlər, nümayişlər və hadisələr qalan faizi təmsil edir.

Le Monde qəzetindən əldə edilən nəticələr göstərir ki, bu qəzetdəki bütün xüsusi isimlərin 50,4%-i sübutlarla müşayiət olunur: bu, kifayət qədər azdır. Şəxs adlarının 93%-i və təşkilat adlarının 65% -i sübut ilə müşayiət olunur. Yer adları çox nadir hallarda sübutla müşayiət olunur (onların yalnız 20%-i). Qalan xüsusi adlar (onları müəyyən etməyə imkan verən aydın ipucları olmadan) buna görə də təhlil edilərək və s. tapılmalı olacaq. Daha ətraflı, bu tip adları çəkilən qurumların hər birinin tanınması üçün hansı vasitələr olacağına nəzər yetirək (nəticələr Le Monde qəzetindən əldə edilmişdir).

Şəxs adları

Antroponimlər (ad və soyadlar) çoxsaylıdır. Buna baxmayaraq, şəxs adlarını müəyyən etmək çox asandır: onları aşkar etmək üçün ipucları çoxdur. Şəxs adlarının 19,4%-i sol kontekstdən istifadə etməklə aşkar edilə bilər (ən çox sivillər, funksiya və ya peşə – *président* prezident, *Mme* xanım, l`*architecte* memar və s.), lakin əksəriyyəti (60,1%) fransız dilində bir ad və ya birmənalı soyad hissəciyi olan daxili sübutla (məsələn: *von, di* və s.) müəyyən edilir. Qeyd edək ki, insanların adlarının 8,6%-i ikiqat sübutla müşayiət olunur: sol kontekst və daxili sübut (məsələn, modelyer Koko Şanel).

Şəxs adlarının 45%-də sivillik, vəzifə və ya peşə adı olan kontekstdən sonra soyad gəlir (*Ouest France* jurnalında bu proporsiya 33% təşkil edir). Buraya bir ad da əlavə edilə bilər (məsələn, *le président français Jacques Chirac* fransa prezidenti Jak Şirak).

Şəxs adlarının 45%-nin təsvir edilə bilən konteksti yoxdur, lakin ad və soyadla təmin edilən daxili sübutu ehtiva edir, məs., *Pierre Vincent.* Adlar morfoloji ipuçları əsasında və lüğətlərdən istifadə etməklə tanınır. Biz adların çıxarılmasına xüsusi əhəmiyyət veririk, çünki adı soyaddan necə fərqləndirəcəyimizi bilsək, şəxs adlarının variantını yaratmaq daha asan olacaq. Tədqiqat işinin əvvəlki bəndlərində göstərdiyimiz kimi şəxs adları və soyadlar, əsasən, sadə və mürəkkəb formalara malik olur. Lakin bu formalardan başqa fransız dilində soyadların fransız formaları (məsələn: *L'Huillier, Le Falch'un*), hissəcikləri olan fransız soyadları (məsələn: *Dupont de Nemours, de Funès, de la Fontaine*).

Şəxs adlarının 5%-nin heç bir konteksti yoxdur, hətta mürəkkəbdirlər ki, bu da onları digər xüsusi isimlərdən mütləq fərqləndirə bilər. Kontekstsiz bu insan adları əsasən çox tanınmış insan adlarıdır ki, mətnin müəllifi məqalədə artıq istinad edilən bu şəxslərin nə adının, nə vəzifəsinin,

nə də peşəsinin qeyd edilməsinə ehtiyac olmadığını düşünür. Məs, *Salvador Dali n'est pas le premier à passer à la postérité commerciale* Salvador Dali kommersiya nəslinə keçən ilk şəxs deyil.

Təşkilat adları

Onomastik vahidlər içərisində çoxlu sayda təşkilat adları var və onlar bir çox variantlara malikdir (məsələn: Organisation du traité de l'Atlantique Nord, traité de l'Atlantique, OTAN – Şimali Atlantika Müqaviləsi Təşkilatı). Bu adlar iqtisadiyyatdan asılı olaraq yaranır və yox olur.

Təşkilat adlarının 51,2%-i daxili sübutla başlayır: təşkilatın adını bildirən böyük hərflə yazılmış ilk söz (məsələn: *Fonds Monétaire International* Beynəlxalq Valyuta Fondu). Onlar, əksər hallarda, təsvir əsası olan xüsusi isimlərdir. Onlar bir cəmiyyət (məsələn: *Société européenne des satellite* Avropa Peyk Cəmiyyəti), təşkilat (məsələn: *Organisation mondiale de la santé* – Ümumdünya Səhiyyə Təşkilatı), bank (məsələn: *Banque d`Europe* Avropa Bankı) və ya digər (məsələn: *Front populaire français* Fransa Xalq Cəbhəsi, *Union cycliste internationale* Beynəlxalq Velosipedçilər İttifaqı) yerlər olduğunu təxmin etməyə imkan verən ümumi isimlərdən ibarətdir.

Daxili sübutları ehtiva edən təşkilatların adlarının 7%-i əslində xarici adlardır (məsələn, *Mellon Foundation* Melon Fondu, *Bank of Australia* Avstraliya Bankı), onları müəyyən etmək olduqca sadədir: onların hər sözünün əvvəlində böyük hərf var, onların tərkibində *of, and* kimi sözlər də ola bilər. Bu xüsusi isimlər sağ tərəfində çox vaxt Ltd, Research (Araşdırma) və s. kimi daxili sübuta malikdir.

1% adlar onların morfologiyası sayəsində tapıla bilər, məsələn, ampersandın (*and* işarəsinin olması) olması (məsələn: *AT&T – American Telefone and Telegraph*).

Sol xarici sübut təşkilat adlarının 12,7%-ni (məsələn: *filiale de Socar*- Sokarın filialı) və sağ kontekst yalnız 1,2%-ni təşkil edir. Xarici sübutla müşayiət olunan təşkilat adları xalis xüsusi isimlər formasını alır: onlar əsasən şirkət adlarıdır. Onların xarici sübutu ümumiyyətlə qrup, agentlik, şirkət və s. kimi bir sözdən və bəlkə də bir toponimik sifətdən (məsələn: *groupe britannique Cable & Wireless* Britaniya Cable & Wireless qrupu, *compagnie Airbus* Airbus şirkəti) ibarət olur.

Yer adları

Toponimlər nisbətən sabit xüsusi adlardır, yəni bu və ya digər toponimə verilən ad nadir hallarda dəyişir (Piton və Maurel 1997), lakin linqvistik dəyişikliklər və xüsusi adların avtomatik tanınması tarixə görə baş verir (*Châlons-sur-Marne* Şalon sür Marn son zamanlarda *Châlons-en-Champagne* Şalon an Şampan adlandırılır). Üstəlik, toponimlərin sayı kifayət qədər məhduddur.

Yer adlarının yalnız 20%-nin sol və bəzən sağ konteksti var (məsələn*: l'estuaire de la Loire* Luaranın mənsəbi, *la mer Noire* Qara dəniz) və bir neçəsinin isə daxili sübutu var. Daxili sübuta malik olanlar əsasən şəhər və ya departament adları (məsələn: *Chaumont-sur-Loire* = tire və "sur (üstündə)" şəhər adları üçün tipikdir) və ya xarici yerlərin adlarıdır (məsələn, *High Street, London park).*

Bütün xüsusi isimlər

Bu mövzuda aparılan araşdırmalar göstərir ki, kateqoriyalara bölünə bilməyən xüsusi isimlər ən azı sintaktik ipucları və böyük hərflərin olması ilə müəyyən edilə bilər. Onları təyin etmək üçün başqa vasitələrdən istifadə etmək lazım gələcək; məsələn, həmin mətndə olan omonimləri ilə eyni tipə aid etmək.

Nəticə. Korpus araşdırması bizə göstərir ki, müxtəlif növ xüsusi adlar müəyyən etmə baxımından bərabər deyildir, çünki onların görünmə kontekstləri və qəzet məqalələrində sayı çox fərqlidir. Buna görə də xüsusi adların müəyyənləşdirilməsi vasitələrini bu fərqlərə uyğunlaşdırmaq lazımdır.

Qaydalara əsaslanan üsullar xüsusi adların böyük bir hissəsini tapmağa imkan verir, lakin onların öz məhdudiyyətləri var: qrammatikanın qaçılmaz natamamlığı bəzi səhvlərin və ya çatışmayan cavabların əsas səbəbidir. Qeyri-müəyyənliklərin hamısı həll olunmur, xüsusi adlar ətrafında kontekstin olmaması onların müəyyənləşdirilməsinə və təsnifləşdirilməsinə mane olur.

Xüsusi isimlərin və onların növlərinin tanınması onun kontekstinin avtomatik və düzgün

tərcüməsini nəzərdən keçirmək üçün böyük maraq doğurur [133, s. 637–650; 280]

ƏDƏBİYYAT SİYAHISI :

1. Daille, B. Reconnaissance automatique des noms propres de la langue écrite: les récentes réalisations. Traitement Automatique des Langues / B.Daille, E. Morin. – Nantes: ATALA/Hermès Science Publications, – 2000. n°41/3. – p.601-621.
2. Friburger, N. Linguistique et reconnaissance automatique des noms propres/N. Friburger // Journal des traducteurs: Meta, Les Presses de l'Université de Montréal, – 2006. 51(4), – p. 637–650.
3. Jacquemin, C., Bush, C. Combining Lexical and Formatting Cues for Named Entity Acquisition from the Web // Dans Proc. Joint SIGDAT Conference on Empirical Methods in NLP and Very Large Corpora. – Hong Kong, – 2000. – p.181-189.
4. Shisha-Halevy, A. The Proper Name: Structural Prolegomena to Its Syntax – A Case Study in Coptic / A. Shisha-Halevy. – Vienne: VWGÖ, – 1989. –143 p.
5. Wolinski, F., Vichot, F. et B. Dillet. Automatic Processing of Proper Names in Texts, Proceedings of the Seventh Conference of the European Chapter of the Association for Computatinal Linguistics (EACL'95) // – Dublin: University College of Dublin, – 1995. – p.23-30.
6. https://doi.org/10.4000/corela.117

SUMMARY

FEATURES OF AUTOMATIC RECOGNITION OF ONOMASTIC UNITS IN FRENCH LINGUISTICS

Vafa Seyid

The article explores the features of automatic recognition of proper nouns in French linguistics, examines the distinguishing characteristics of proper nouns from common nouns, and analyzes various linguistic scholars' views on the existence of a proper noun as a separate individual or object.

It is noted that the recognition of proper nouns within the text sometimes presents challenges for linguists. From the perspective of automatic language processing, the work on the recognition of proper nouns compels computer scientists to propose simpler typologies and practical solutions for computer work, while still adequately considering the reality of proper nouns.

It is proposed to distinguish three types of objects for the extraction, recognition, and categorization of named entities: ENAMEX, TIMEX, and NUMEX. Three main system types are presented for distinguishing proper nouns: rule-based systems, learning systems, and hybrid systems.

The article demonstrates that the most reliable clue for detecting and classifying proper nouns is their left or right context and internal structure. Proper nouns can also function as epithets, attributes, subjects, objects, or modifiers in sentence constructions.

The research shows that different types of proper nouns are not equal in terms of identification, as their contextual appearance and frequency in newspaper articles vary significantly. Therefore, the tools for identifying proper nouns must be adapted to these differences.

Keywords: *automatic identification, hybrid systems, delimitation, variation, internal and external evidence, context, category.*

РЕЗЮМЕ

ОСОБЕННОСТИ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ОНОМАСТИЧЕСКИХ ЕДИНИЦ В ФРАНЦУЗСКОМ ЯЗЫКОЗНАНИИ

Вафа Сейид

Статья исследует особенности автоматического распознавания собственных существительных в французском языкознании, рассматривает отличительные характеристики собственных существительных по сравнению с нарицательными и анализирует различные мнения

лингвистов о существовании собственного существительного как отдельного индивида или объекта.

Отмечается, что распознавание собственных существительных в тексте иногда вызывает трудности у лингвистов. С точки зрения автоматической обработки языка, работа по распознаванию собственных существительных заставляет специалистов по вычислениям предлагать более простые типологии и практичные решения для работы с компьютерами, при этом адекватно учитывая реальность собственных существительных.

Предлагается различать три типа объектов для извлечения, распознавания и категоризации именованных сущностей: ENAMEX, TIMEX и NUMEX. Три основных типа систем представлены для различения собственных существительных: системы на основе правил, системы обучения и гибридные системы.

Статья демонстрирует, что наиболее надежным признаком для обнаружения и классификации собственных существительных является их левый или правый контекст и внутренняя структура. Собственные существительные также могут функционировать как эпитеты, атрибуты, субъекты, объекты или модификаторы в конструкциях предложений.

Исследование показывает, что различные типы собственных существительных не равны с точки зрения идентификации, поскольку их контекстуальное появление и частота в газетных статьях значительно различаются. Поэтому инструменты для идентификации собственных существительных должны быть адаптированы к этим различиям.

Ключевые слова: *автоматическая идентификация, гибридные системы, делимитация, вариация, внутренние и внешние доказательства, контекст, категория*.