Obi-ugor morfológiai elemzők és korpuszok

Fejes László[1], Novák Attila[2]

[1]MTA Nyelvtudományi Intézet
1068 Budapest, Benczúr utca 33.
fejes@nytud.hu
[2]MorphoLogic
1116 Budapest, Kardhegy utca 5.
novak@morphologic.hu

Kivonat: Cikkünkben a végéhez közeledő OTKA NF 71707 projekt keretein belül létrehozott obi-ugor számítógépes morfológiákat, annotált korpuszokat, a használatukat lehetővé tevő webfelületet és azokat a problémákat mutatjuk be, amelyek a fejlesztés során felmerültek.

1 Bevezetés

A kisebb uráli nyelvek veszélyeztetettek, ezért dokumentálásuk nemzetközi jelentőségű feladat. A magyarországi uralisztika ezen a területen jelentős hagyományokkal rendelkezik: a 19. század közepétől kezdve magyar kutatók rendszeresen gyűjtöttek szövegeket, szótári anyagokat, és ezek alapján készítettek grammatikai vázlatokat is. Végéhez közeledő projektünkben (OTKA NF 71707) a korábban gyűjtött obi-ugor szövegek számítógépes feldolgozásával morfológiailag annotált korpuszokat hoztunk létre.

A projekt a két obi-ugor nyelv három nyelvjárását öleli fel, és az alábbi négy fő modulra oszlik:

Vogul (manysi) északi nyelvjárás: Kálmán Béla gyűjtése (WT) [6]
Vogul (manysi) északi nyelvjárás: Munkácsi Bernát gyűjtése (VNGY) [7]
Osztják (hanti) szinjai nyelvjárás: Ruttkay-Miklián Eszter gyűjtése
Osztják (hanti) kazimi nyelvjárás: különböző gyűjtések [12, 15, 14]

A modulokban egy-egy gyűjtés, illetve nyelvjárás feldolgozására vállalkoztunk. Ezt az indokolja, hogy a számítógépes elemzés megköveteli a lehető legegységesebb korpuszok használatát: a sokszínű korpuszokhoz megengedőbb elemzőt kellene építeni, ami viszont óhatatlanul a téves elemzések megszaporodásával járna együtt. Éppen ezért minden egyes tér- és időbeli nyelvváltozathoz önálló elemzőt építettünk.

Az elkészült elemzők és korpuszok egy része már online hozzáférhető, és folyamatosan tesszük közzé az újabb elkészült erőforrásokat [16].

2 Az elemzők építése

A hanti nyelvjárások közötti igen jelentős különbségek miatt a két hanti elemzőt egymástól függetlenül, az alapoktól építettük fel. A két manysi gyűjtés esetében ugyanazon nyelvjárás két időben eltérő nyelvállapotát két igen eltérő transzkripcióval rögzítették: ez indokolta, hogy itt is két külön morfológiát hoztunk létre. A Kálmán jelen projekt keretében feldolgozott szövegeihez készült elemző esetében támaszkodhattunk egy korábbi projekt keretében Kálmán által máshol [5] publikált szövegekhez készült elemzőnkre. Munkácsi szövegei esetében azonban ismét az alapoktól kellett kezdenünk a munkát.

A manysi elemzők tőtárát az adott kiadványokhoz készült szójegyzék [6], illetve szótár [9] alapján készítettük el. A hanti tőtárak alapjául elsősorban Steinitz szótára [13] szolgált. A szövegek feldolgozása során az egyik legfőbb problémát a szövegek belső inkonzisztenciája és a szótárak pontatlansága okozta. A másik probléma a nyelvtanok ([5, 6, 8, 12, 13]) vázlatszerűsége és felületessége volt: ezek ritkán adtak elég támpontot a morfofonológiai jelenségeknek a számítógépes implementációhoz szükséges pontos leírásához. Cikkünkben bővebben kitérünk néhány olyan nyelvtani problémára, amelyek megoldása jelentős kihívást jelentett.

A tőtárak és a szövegek nagy részének digitalizálása begépelés útján történt, a Munkácsi–Kálmán szótárat [9] pedig (ez a Munkácsi által gyűjtött és publikált szövegek szóanyagát fedi le) OCR-rel digitalizáltuk. A Munkácsi–Kálmán szótárban alkalmazott manysi átírás számtalan szokatlan karaktert tartalmaz (magánhangzóbetűket több különböző ékezettel, felső indexben álló gammákat stb.), ezért az OCR programot egyedileg kellett betanítani a feladatra. Ráadásul a szótárban szereplő dőlt betűs cirill karakterek egy része (*a, c, e, m, n, o, p, x, y*) megkülönböztethetetlen a manysi címszavakban álló dőlt betűs latin karakterektől, ezért ennek a megkülönböztetésnek a felismerését nem bíztuk az OCR programra, hanem az összes ilyen karaktert cirillként ismertettük fel a programmal, és utólag automatikusan konvertáltuk manysi részekben álló karaktereket. Konverzió után az OCR-hibákat kézzel javítottuk. A szótárban a tipográfia alapján programmal azonosítottuk a címszavakat és a magyar, német, illetve helyenként orosz nyelvű fordításokat, a nyelvjárásra vonatkozó adatokat, így képezte a szótár a Munkácsi-szövegek feldolgozására készülő manysi elemző tőtárának alapját. A szótár és Munkácsi szövegkiadásai más manysi nyelvjárások szóanyagát is tartalmazzák. Jelen projektben azonban csak a legbővebben adatolt északi nyelvjárás feldolgozására vállalkoztunk.

3 A morfológiai elemzők jellemzői

A projekt keretében elkészült morfológiai elemzők mindegyike a MorphoLogic *Humor* elemzőmotorjára épül. A morfológiai adatbázisok létrehozására a korábban már számos más nyelv (elsőként a magyar) számítógépes morfológiájának létrehozásához használt morfológiaiadatbázis-leíró keretrendszert használtuk ([10, 11]). A Humor elemző morfémaallomorfok felszíni alakjainak egy véges állapotú automata által leírt szónyelvtannak és a lokális szomszédossági megszorításoknak is megfelelő sorozatait

ismeri fel a bemenetén kapott szóalakban, és az ezeknek megfelelő morfémasorozatokat jeleníti meg elemzésként. A rendszert kiegészítettük egy olyan mechanizmussal, amely az eredetileg a morfológia forrástőtárában tárolt különböző nyelvű glosszákat is az elemzésekhez csatolja, így a rendszer egyben szemantikai címkézést is végez. Az elemzőkhöz készített webes felületen így az elemzések magyar és angol (illetve a manysi elemzők esetében emellett még német) nyelvű glosszákkal együtt jelennek meg. Ez lehetővé teszi, hogy a szövegeket a nyelvet nem beszélő kutatók is értelmezni, illetve egyértelműsíteni tudják.

A keretrendszerben a tövek és a toldalékok leírására különböző formalizmus szolgál, de mindkettőben általában csak morfémák és megjósolhatatlan lexikai jegyek redundanciamentes leírása szerepel. Az elemző által használt allomorfokat és a morfok szomszédossági megszorításait leíró teljes jegyegyütteseket az elemző lexikonának kompilálásakor a keretrendszer állítja elő a morfológia forrásának részét képező szabályrendszer felhasználásával. A forráslexikonban allomorfok, illetve toldalékolt alakok csak akkor szerepelnek, ha olyan mértékben rendhagyóak, hogy szabállyal való előállításuknak nem láttuk értelmét.

A jelen projekt keretében feldolgozott nyelvek és korpuszok esetében azonban jóval gyakoribb eset volt, hogy a lexikonba allomorfokat, írásváltozatokat kellett felvennünk, mint például a sztenderd mai magyar szövegek elemzésére készített elemzőnk esetében, mert itt nagyságrendekkel több a lejegyzési következetlenség, illetve a nyelvek kevésbé sztenderdizált voltából adódóan is jóval nagyobb a változatosság. Az alábbi táblázat ezt szemlélteti.

elemző	lexikálisan megadott allomorffal vagy toldalékolt alakkal rendelkező tövek aránya	
mai magyar:	274/139859	0.20%
manysi WT	475/4209	11.29%
manysi VNGY	3705/16526	22.42%
szinjai hanti	314/2606	12.05%
kazimi hanti	301/1958	15.37%

A következő táblázatban összefoglaltuk az egyes nyelveken, nyelvváltozatokon rendelkezésünkre álló, illetve feldolgozott korpuszok és az elkészült elemzők mennyiségi jellemzőit.

Nyelv	korpusz	tőlexikon				toldaléklexikon	
	szó	lemma (*jelentés)		allomorf		mögöttes alak	allomorf(sorozat)
		zárt	nyílt	zárt	nyílt		
manysi WT	10659	387	3822	622	5483	376	5285
manysi VNGY	81717 (1026)	909	15617	1900	34665	297	2944
szinjai hanti	151500 (6539)	256	2350	615	7894	140	813
kazimi hanti	19228	209	1749	689	6756	150	1491

A táblázatban külön oszlopban soroltuk fel a nyílt (főnév, melléknév, ige, határozószó) és a zárt szófajosztályokba (többi szófaji kategória) tartozó tövek számát. A tövek elkülönült jelentései külön tételként jelennek meg a tőtárakban. A táblázatból kitűnik, hogy a hanti elemzők esetében az egyes morfémáknak átlagosan több mint 3 allomorfja van, ami az alább részletezett szótagszerkezeti megszorításokból, az azok megvalósítására a beszélők által alkalmazott stratégiák változatosságából, valamint a lejegyzésekben tapasztalható ingadozásából adódik. A megadott korpuszméreteknél néhol szereplő zárójeles szám egy olyan alaposabban ellenőrzött részkorpusz méretére utal, amelyeken belül igyekeztünk minden lejegyzési hibát kijavítani, és az elemző által teljes lefedést biztosítani.

A toldaléklexikonokban toldalékkapcsolatok is szerepelnek, illetve az inflexióstoldalék-sorozatok nagy részét a keretrendszer offline kigenerálja, így az elemző gyorsabban működik, mert a teljes sorozatot egy lépésben találja meg elemzéskor a lexikonában. Ebből adódik a mögöttes toldalékok és az elemző kigenerált allomorflexikonjának mérete közötti sokszoros különbség.

4 A morfológiai elemzők jelentősége

A morfológiai elemzők használatával előállítható, morfológiailag annotált korpuszok jelentőségéről itt nem kívánunk szólni, ezek haszna minden szakmabeli számára nyilvánvaló. Azt azonban jeleznünk kell, hogy a projekt sajnos még nem foglalta magában egy komplex korpuszkezelő fejlesztését, így az ilyesféle lehetőségek – például kifinomult keresőrendszer hiányában – korlátozottak.

Fontosnak érezzük azonban szólni a morfológiai elemző fejlesztése során nyert tapasztalatok jelentőségéről.

Az obi-ugor nyelvek kutatásának lehetőségei – bár ma is élő nyelvekről van szó – nagyjából a holt nyelvek kutatásának lehetőségeihez hasonlíthatók. Élő nyelvhez hasonlóan csak az éppen terepen levő nyelvész kutathatja, ilyen jellegű munkára azonban ritkán nyílik alkalom, s mivel a terepmunkás is tisztában van az alkalom különleges voltával, idejét leginkább nyelvi anyag (szövegek) rögzítésére fordítja. Maga a nyelvészeti kutatás elsősorban ezekre a szövegekre épül, azaz az obi-ugor nyelvészet szorosan összefonódik az obi-ugor filológiával. Mivel egy-egy nyelvjárásról, illetve annak időbeli állapotáról mindig igen korlátozott adatunk van, és az egyik nyelvjárásban vagy állapotban megfigyelt szabályszerűségeket nem vetíthetjük át automatikusan más nyelvjárásokra és állapotokra, az adatok kezelése nagy óvatosságot és pontosságot igényel.

A számítástechnika előtti korszakban az adatok gyűjtése, kezelése, feldolgozása rengeteg hibalehetőséget rejtett magában. Nem csupán az adatok rögzítésekor kerülhetett hiba a rendszerbe, az adatokat is kézzel másolták, a sajátos jelek kezelése a nyomda számára is nehézséget jelentett. A hibákat nehéz volt kiszűrni, hiszen a kiadott szövegekben, a szótárakban és a nyelvtanokban szereplő adatok többé nem „találkoztak” egymással. Egy lexikai jellegű tanulmány már nyilvánvalóan a szótárra épült, nem ment vissza a szövegekhez. Azok a hibák, melyek a szövegek feldolgozásakor és a szótár készítésekor keletkeztek, torzították a nyelvről alkotott képet.

A számítógépes morfológiai elemzők nagy előnye, hogy a korpuszban és a tőtárakban levő adatok, illetve az explicit módon, képletszerűen megfogalmazott morfofonetikai és morfológiai szabályok interakcióban vannak egymással, a közöttük levő ellentmondások az esetek nagy részében szükségszerűen nyilvánvalóvá válnak.

Az általunk épített manysi elemzőkben mindig az adott szövegkorpuszhoz kiadott szójegyzékeket, illetve szótárat használtuk. Mindhárom esetben kiderült, hogy a szójegyzékek, illetve a szótár hibásak, illetve hiányosak. A szavak nem ugyanabban az alakban szerepelnek a szótárban, mint a szövegekben (jellemző például a magánhangzók hosszúságának eltérő jelölése, de gyakori a puszta helyesírási következetlenség, pl. a kötőjel használatában való ingadozás is), vagy nem szerepel a szövegben előforduló összes alakváltozat. Egyes szavak teljesen hiányoznak, különösen gyakori ez a képzett szavak esetében (olyanoknál, melyek alapszava szótározva van), illetve a tulajdonneveknél. Vannak esetek, amikor a szótár szerint a szó nem dokumentált az általunk vizsgált északi nyelvjárásban, szövegeinkben azonban mégis szerepel. Az összetett szavak szótározása is meglehetősen rapszodikus: egyes transzparens összetételek szerepelnek a szótárban, miközben sajátos jelentésű összetételek hiányoznak. Az elemzők fejlesztése során véletlenül bukkanunk olyan esetekre, amikor a szó ugyan szótározva van, de nem minden, a szövegekben dokumentált jelentésében. Az ilyen esetek módszeres felderítésére majd a teljes korpuszok egyértelműsítése fog lehetőséget teremteni.

A hanti korpuszok esetében a feldolgozott szövegekhez nem készültek szójegyzékek, ezeket mi magunk hozzuk létre. A Steinitz-féle szótárral [13] való egybevetés ugyan fontos szerepet játszott, de mindkét korpuszunkban jócskán találtunk olyan töveket, melyek Steinitznél nem, vagy más alakban szerepeltek.

Pusztán az a tény, hogy a szövegek digitalizálva vannak, lehetőséget teremt a lejegyzés egyenetlenségeinek korrigálására. Így például az alakváltozatok megjelenésének aránya utalhat arra, hogy mikor lehet szó valódi alakváltozatokról, és mikor valószínűbb, hogy egyes írott „alakváltozatok” csupán sajtóhiba eredményei. A szöveg feldolgozásának később stádiumában más lejegyzési egyenetlenségek kiküszöbölésére is sor kerülhet, így például a hol külön, hol összetett szóként leírt szószekvenciák lejegyzése egységesíthető. A szövegek digitalizálásának köszönhető, hogy felfedeztük: a Munkácsi–Kálmán szótárban [9] olyan szóalakok is szerepelnek, amelyek a szövegben [7] nem – ezek feltehetően Munkácsi kéziratos cédulaanyagából kerültek a szótárba. Ennél azonban sokkal érdekesebb, hogy a szótárban olyan példamondatok is vannak, melyek a kiadott szövegekben nem lelhetők fel. Ennek alapján azt gyanítjuk, hogy Munkácsi cédulái jelentős korpuszt, ha nem is szövegeket, de elszigetelt mondatokat tartalmaz. Okkal feltételezhetjük, hogy ezen példamondatoknak töredéke került csak be a szótárba. Sikerült tehát (újra)felfedeznünk egy olyan 19. századi manysi forrásanyagot, mely időközben kiesett a kutatás látóköréből, és a morfológiai elemzés fejlesztése nélkül talán örökké „elveszett” volna. A cédulaanyag ilyen típusú feldolgozására egy további projekt folyamán kerülhet sor, mindenesetre ezt is feladataink között tarjuk számon.

A morfológiai elemző építése során erősen támaszkodtunk a szóban forgó nyelvjárásokat leíró nyelvtani vázlatokra. Ezek – érthető módon – nem olyan egzakt leírásokat tartalmaznak, melyek azonnal alkalmasak szabályokba kódolásra, de mindenesetre jó kiindulópontul szolgálnak. A morfofonológiai váltakozások közül az obi-ugor

nyelvekben a legjelentősebbek a jól formált szótagok építését célzó szabályok. Ezt minden nyelvváltozatban vegyes stratégiával érik el: részben mássalhangzók törlésével, részben magánhangzók (elsősorban svá) betoldásával. A helyzetet nagyban bonyolítja, hogy a szonoránsok előtti svábetoldás helyett gyakran a szonoráns válik szótagalkotóvá – legalábbis a lejegyzésben ez szerepel. Vannak azonban helyzetek, amikor a lejegyzés sem a svá betoldását, sem a szonoráns szótagalkotóvá válását nem jelzi. Kezdetben azt feltételeztük, hogy ezekben az esetekben egyszerűen a lejegyzés pontatlanságáról van szó.

Későbbi megfigyeléseink azonban ezt megkérdőjelezik. Nem egy esetben svá előtt a tőnek az az alakváltozata jelenik meg, amelynek szabályszerűen magánhangzóval kezdődő toldalék előtt nem lenne szabad megjelennie. Úgy tűnik, a rosszul formált szótagszerkezet kiküszöbölésére két szabály is aktivizálódik, holott az egyik bőven elegendő lenne. Ennek megfelelően az elemzőben azokat a toldalékokat, amelyek szonoránssal kezdődnek, akár svá-betoldásos alakjukban, akár svá nélkül jelennek meg, sem magán-, sem mássalhangzós kezdetűként nem jelöljük meg, így mindkét tőalakváltozathoz kapcsolódhatnak.

Más esetekben viszont nem toldódik be svá, a tőnek mégis az az alakváltozata jelenik meg, amelyet csak magánhangzós toldalékok előtt várnánk. Nehéz eldönteni, hogy ilyen esetekben nem egyszerű sajtóhibáról van-e szó. Amióta azonban felfigyeltünk a problémára, több független forrást is felfedeztünk, melyek azt a benyomásunkat erősítik meg, hogy ez igenis előfordulhat. Pillanatnyilag azt a megoldást követjük, hogy a mássalhangzóval kezdődő toldalékok előtti mássalhangzókapcsolat-egyszerűsödések fakultatívak: az elemzéskor nem várjuk el a svá-betoldást, ám a szóalak-generátor a svát mindig betoldja.

Elképzelhető azonban, hogy szabályaink túlságosan megengedőek. Előfordulhat például, hogy az általunk homonimként kezelt toldalékok a morfofonológiai váltakozásokban eltérő viselkedést mutatnak. Ezt azonban csak a kutatás egy későbbi szakaszában, az egyértelműsítés elvégzése után lehet vizsgálni: az, hogy valójában melyik morfémának milyen allomorfjai jelenhetnek meg a különböző környezetekben, csak a már egyértelműsített szövegeken vizsgálható. Ám ekkor sem lesz könnyű elkülöníteni a sajtóhibákat a valódi alternánsoktól.

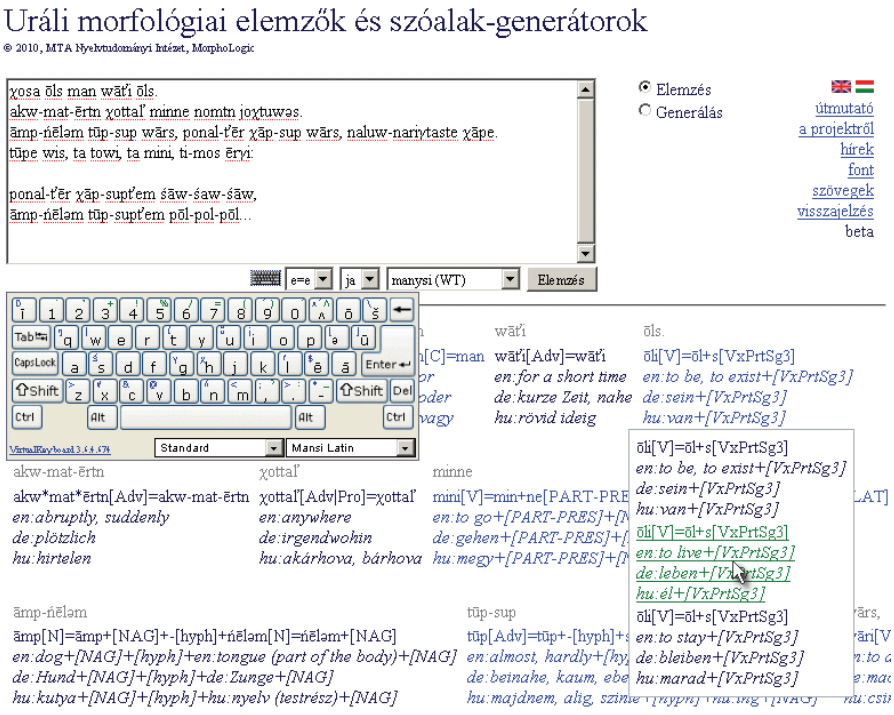
Az első obi-ugor elemző készítése során elsősorban a nem első szótagban található magánhangzók minősége, illetve a svá betoldása és be nem toldása kapcsán vetett fel kérdéseket. A problémák megoldása céljából több kutatás indult el, köztük akusztikai vizsgálatok is: ezekről több előadás és cikk is született ([1, 2, 3, 4]). A jelenlegi problémák inkább fonológiaelméleti kérdéseket állítanak a központba: hogyan lehetséges az, hogy miközben egy nyelv radikális váltakozásokat vezet be a rosszul formált szótagok kiküszöbölésére, ezzel egy időben nagyfokú toleranciát is mutat ezen rosszul formált alakokkal szemben. E kérdéssel kapcsolatban is újabb tanulmányok sora várható.

5 Online morfológiák

A projekt keretében készült morfológiák és a korpuszt alkotó szövegek a projekt végére webes felületen keresztül válnak elérhetővé [16]. Az elemzők esetében a kivá-

lasztott szöveget a megfelelő ablakba másolva a felhasználó megkapja a szövegben szereplő szavak lehetséges morfológiai elemzéseit és az elemzésekben szereplő tőmorfémák jelentését. Virtuális billentyűzet segítségével maga is gépelhet be szövegeket. Az elemzéseket megjelenítő webes felület egyben kézi egyértelműsítő eszközként is szolgál: a többértelmű szavak elemzései pop-up ablakban jelennek meg, ha az egeret egy többértelmű szó fölé mozgatjuk, ezek közül egérrel választhatunk. Az elkészült elemzések, illetve azok egyértelműsített változata elmenthető, az elmentett változatot a böngészőbe betöltve, az esetlegesen félbehagyott egyértelműsítő munka később folytatható.

A webes felületen keresztül nemcsak morfológiai elemzők, hanem szóalak-generátorok is elérhetők az egyes nyelvekhez. Az alábbi képernyőképek illusztrálják a szövegbeírás, a virtuális billentyűzet, az egyértelműsítő felület és a szóalak-generátor használatát. Ha egy adott morfémasorozat több formában is megjelenhet, akkor a generátor kimenete az elemző többértelmű kimenetének megjelenítéséhez hasonlóan jelenik meg a webes felületen, a lehetséges szóalakváltozatok itt is az egérmutatót a generált szóalak fölé mozgatva megjelenő pop-up ablakban láthatóak.



Bibliográfia

1. Bakró-Nagy M., Fejes L.: Schwa or not schwa? Synchronic and diachronic speculations on an Ob-Ugric vowel. FUSAC, Vancouver. 2008. június 8.
2. Fejes L.: A vogul morfológiai elemző(k) felé. Fonológiai és morfológiai megfigyelések. Obi-ugorok a 21. században (CD-ROM). MTA Nyelvtudományi Intézet, Budapest (2006) http://fgroszt.nytud.hu/publikaciok/obi-ugorok/text/nyelv2.html
3. Fejes L.: Az északi-manysi vokalizmus néhány kérdése. MTA Nyelvtudományi Intézet, 2008. május 8. http://nytud.hu/~fejes/pdf/manysiV/manysi_V-k_ea.pdf
4. Fejes L.: On the acoustics of the Northern Mansi Vowel System. Poszterelőadás a 17. Manchesteri Fonológiai Találkozón. 2009. május 29. http://fgrtort.nytud.hu/images/stories/fejes/fejes_manchester_poster.pdf
5. Kálmán B.: Chrestomathia Vogulica. Tankönyvkiadó, Budapest (1989)
6. Kálmán B.: Wogulische Texte mit einem Glossar. Akadémiai Kiadó, Budapest (1976)
7. Munkácsi B.: Vogul népköltési gyűjtemény. 1–4. Budapest (1892–1921)
8. Munkácsi B.: A vogul nyelvjárások szóragozásukban ismertetve. Budapest (1894)
9. Munkácsi B., Kálmán B.: Wogulisches Wörterbuch. Akadémiai Kiadó, Budapest (1986)
10. Novák A.: Milyen a jó Humor? In: Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2003). Szegedi Tudományegyetem, Szeged (2003) 138–145
11. Prószéky G., Novák A.: Computational Morphologies for Small Uralic Languages. In: Arppe, A., Carlson, L., Lindén, K., Piitulainen, J., Suominen, M., Vainio, M., Westerlund, H,. Yli-Jyrä, A. (szerk.): Inquiries into Words, Constraints and Contexts. Festschrift in the Honour of Kimmo Koskenniemi on his 60th Birthday. Gummerus Printing, Saarijärvi/CSLI Publications, Stanford (2005) 116–125
12. Rédei K.: Nord-ostjakische Texte (Kazym Dialekt) mit Skizze der Grammatik. Vandenhoeck and Ruprecht, Göttingen (1968)
13. Steinitz, W.: Dialektologisches und Etymologisches Wörterbuch des Ostjakischen Sprache. Akademie-Verlag, Berlin (1966)
14. Steinitz, W.: Ostjakologische Arbeiten. Beiträge zur Sprachwissenschaft und Ethnographie. Herausgegeben von Gert Sauer und Renate Steinitz. Bd. I–IV. Akadémiai Kiadó – Akademie-Verlag, Budapest – Berlin (1980)
15. Хомляк, Л. Р. (ред.): Арєм-моньщєм ел ки мăнл... (Если моя сказка-песня дальше пойдёт...) Фольклорное творчество Пелагеи Алексеевны Гришкиной из деревни Тугияны. ГРУПП »Полиграфист«, Ханты-Мансийск (2002)
16. http://www.morphologic.hu/urali/index.php