

FEJES LÁSZLÓ

## A morfok ábrázolásai a permi nyelvek helyesírásában

The article describes the pure orthographical alternations of the radical and affixal morphs in Komi(-Zyryan) and Udmurt. The author lists the orthographical types of roots and affixes and gives a table of the possible combinations of the forms.

A néhány száz ezer fő által beszélt finnugor nyelvek számítógépes morfológiai vizsgálatánál szembe kell néznünk azzal a ténnyel, hogy korpuszunkba leginkább olyan szövegeket tudunk beszerezni, melyek a mai, cirill alapú helyesírással íródtak. Ezekben a szövegekben az azonos alakú morfok is különböző betűsorokként jelennek meg, attól függően, hogy milyen más morfok állnak a szomszédságukban. A különböző váltakozások leírása elengedhetetlen ahhoz, hogy az egyes morfokat azonosíthassuk.

Közismert tény, hogy az oroszországi nemzetiségek írásbeliségének kialakulására erősen hatott az orosz írás: nem kizárólag a cirill ábécét vették át, de a mássalhangzók lágyságának és a [j] hangnak sajátos magánhangzójelekkel (*я, е, ё, ю* és részben az *у*) való jelölését is. Ezeknek a nyelveknek a fonetikai–fonológiai–morfofonológiai jellegzetességei eltérnek az oroszéitól. Egyfelől be kellett vezetni néhány speciális betűt, melyek az oroszból hiányzó fonémákat jelölték; másfelől sajátos jelölési módokat kellett bevezetni annak érdekében, hogy a lágyítást vagy annak elmaradását olyan esetekben is jelölni tudják, amelyek az oroszban nem fordulnak elő. A fonémák egyértelmű jelölését így sem sikerült minden esetben elérni.

A két permi nyelv ilyen szempontból szerencsésnek mondható, mert helyesírásuk alkalmas minden fonémájuk jelölésére. Magánhangzórendszerükben egy olyan fonéma van, amelyhez hasonló az oroszban nincs: a középső nyelvállású illabiális centrális (ennek jele mindkét nyelv helyesírásában *ö*). Önálló fonéma a felső nyelvállású illabiális centrális, mely artikulációjában közel áll az orosz */i/* kemény mássalhangzók után megjelenő allofónjához, így betűjele is azéval egyezik meg: *ы*. Ez a két fonéma megjelenik palatális mássalhangzók után is, ilyenkor a palatilizáltságot lágyságjel, a [j]-t a cirill *й* jelöli. A mássalhangzórendszerben három olyan fonémát találunk, amely az oroszból hiányzik: a nem palatális zöngétlen affrikáta, a nem palatális zöngés affrikáta, ill. a palatális zöngés affrikáta. Ezeket az udmurtban mellékjeles betűk (*ӱ, ӱӓ, ӱӓ*), a komiban betűkapcsolatok (*тӱӱ, ӱӓӓ, ӱӓӓ*) jelölik. Ezen kívül szükség volt az olyan [i] jelölésére,

amely palatális párral rendelkező nem palatális mássalhangzó után áll. Ennek jelölésére az udmurt a *ü*, a komi a *i* betűt használja. A helyzetet bonyolítja, hogy nem minden palatális hangot jelöl a nem palatális pár és a palatalizációt is jelölő magánhangzójel vagy a lágyságjel kapcsolata, hiszen a palatális affrikátákat az udmurtban a *ч* és a *ӧ* betűk, a komiban a *ч* betű és a *ɔʒ* betűkapcsolat jelöli. Ráadásul az orosz jövevényszók megőrzik eredeti helyesírásukat: a *район* 'járás, kerület' szóban a [jo] hangkapcsolatot *ӧo* jelöli, ami eredeti komi vagy udmurt szavakban nem fordulhatna elő; a *словарь* 'szótár' szóban *рь* van, holott palatális [ř] egyik irodalmi nyelvben sincs. Láthatjuk tehát, hogy bár minden fonéma jelölhető, a jelölési módok gyakran igen körülményesek.

A következőkben azt szeretném megvizsgálni, hogy az egyes morfoknak<sup>1</sup> milyen írott alakjaik vannak, ill. ezek milyen helyzetekben jelennek meg a szavakban. Hasonló vizsgálatról nincs tudomásom, pedig egy ilyen leírás feltétlenül szükséges, ha az adott nyelvekre automatikus helyesírás-ellenőrzőt szeretnénk készíteni. A következőkben a morfok írásbeli alapalakjának azt a formát érdemes tekinteni, ahogyan az adott hangsort önálló használatban leírnánk (függetlenül attól, hogy az adott alak a valóságban előfordulhat-e önállóan). A következőkben az elől (balra) álló elemet tőnek, az őt követő elemet toldaléknak fogjuk nevezni, de ugyanakkor fontos megjegyezni, hogy ugyanezek a szabályok érvényesek a toldalékok, ill. az egy szóba írt összetételi tagok határán is.

## 1. A toldaléktípusok

Nem kell számolnunk *ʋ*, *ʋ*, továbbá a komi *i*, udmurt *ü* toldalékezdetekkel, hiszen ezek önálló szavak elején sem fordulnak elő. A *φ*, *x*, *у*, *у* betűk csak jövevénytövekben fordulnak elő, toldalékokban soha, az idegen utótagú összetételeket pedig nem szokás egybeírni, így az ilyen esetekkel valószínűleg nem kell számolnunk.

### 1.0. Az alternációban részt nem vevő toldalékok

A következő betűkkel kezdődő toldalékok sem maguk nem változnak, sem az előttük álló tőben nem váltanak ki váltokozást: *б*, *в*, *г*, *д*, *ж*, *з*, *ӱ*<sup>2</sup>, *к*, *л*, *м*, *н*, *о*, *п*, *с*, *т*, *ч*, *ш*, *ы*, továbbá az udmurtban a *ӱ*, *ӱч*, *ӱ*: összefoglalva a mássalhangzó-

<sup>1</sup> Morfokról van szó, hiszen az egyes morfémák különböző allomorfjai a helyesírásban különbözőképpen viselkedhetnek, és minden egyes allomorfnak több alakja lehet a helyesírásban. Ugyanakkor hiba lenne az allomorfok írásbeli változatairól beszélni, hiszen az egyes alakváltozatok attól függetlenül alakulnak, hogy milyen funkciója van az adott szegmentumnak.

<sup>2</sup> Csak a *-ӱш*, *-ӱө* és *-ӱы* toldalékezdet jöhet számításba.

kat és a centrális magánhangzókat jelölő betűk tartoznak ide. Mivel ezek a toldalékok nem vesznek részt a váltakozásban, a nulladik típusba soroljuk őket.

### 1.1. Az *a*, *o* és *y* kezdetű toldalékok

Az ilyen toldalékok kezdőbetűje a *я*, *ë*, *ю* kezdőbetűkkel váltakozik. A másodlagos alak akkor jelenik meg, ha a tő lágyságjelre vagy *ÿ*-re végződik: ilyenkor ezek kiesnek. Pl. udm. *пунь* + *o*: *пунĕ* 'fog-as'; *вöü* + *o*: *вöĕ* 'vaj-as'. Ezeket a toldalékokat az egyes típusba soroljuk.

### 1.2. A *я*, *e*, *ë* és *ю* kezdetű toldalékok

Ha ezek a toldalékok mássalhangzót jelölő betűre végződő tövek után kerülnek, akkor előttük keményjel jelenik meg. Sajátos eset a komiban, hogy a *ч* és a *ðз* után a keményjel helyére lágyságjel kerül. Pl. udm. *мон* + *я*: *монья* 'szerintem (én-szerint)'. Ezt nevezzük kettes típusnak.

### 1.3. Az *u* kezdetű toldalékok

Amikor az ilyen toldalék lágyságjelre végződő tőhöz kapcsolódik, a lágyságjel kiesik. Pl. udm. *мадь* + *уз*: *мадюз* 'mesél-t'. Ha viszont olyan tőhöz járul, melynek végén lágy párral rendelkező kemény mássalhangzó (*ð*, *з*, *л*, *н*, *с*, *т*<sup>3</sup>) áll, akkor az *u* helyett a komiban *i*, az udmurtban *ÿ* jelenik meg. Pl. komi *сөм* + *u*: *сөми* 'ad-tam', udm. *сөм* + *u*: *сөмÿ* 'ua.'. Ez lesz a hármas toldaléktípus.

### 1.4. Az *ə* kezdetű toldalékok

Amikor lágyságjelre vagy *ÿ*-re végződő tövekhez járul, a lágyságjel (ill. a *ÿ*) kiesik, a toldalékkezdő betű pedig *e*-vé válik. Pl. udm. *узь* + *ə*: *узе* 'alsz-ik', *ваÿ* + *ə*: *вае* 'hoz-VxSg3'. Akkor is *e*-vé válik, ha lágy párral nem rendelkező kemény mássalhangzó, továbbá *ч* és a komi *ðз*, *ðж* udmurt *ž*, *žc* után kerül – ekkor a tő nem változik. Pl. udm. *сәнез* + *ə*: *сәнезе* 'csizmám'. Ez a típus a négyes.

<sup>3</sup> Nem tartoznak ide az affrikáták, melyek lágyságát másképp jelöljük.

## 2. A tőtípusok

Tő végén csak egyetlen betű nem fordulhat elő: a keményjel.

### 2.1. A magánhangzóra végződő tövek

A magánhangzós tövek nem alternálnak és nem váltanak ki alternációt. Megjegyzendő azonban, hogy az udmurtban gyakori a hiátustöltés, ill. a többesszám jelének (-jos/-os) a [j] nélküli változata jelenik meg, így az ilyen tövek után megjelenő toldalékok írásképe nem egységes: változataik azonban nem helyesírás, hanem morfofonológiai okokra vezethető vissza. Ezt a tőtípust nevezzük A típusnak.

### 2.2. A lágy párral nem rendelkező kemény mássalhangzóra végződő tövek

Ide tartoznak a nem palatális affrikáták (komi *mu*, *ɖʃ*; udmurt *ü*, *ɨc*) is, ezek palatális párját ugyanis nem lágyságjellel és lágyító magánhangzójelekkel jelöljük. Ezek a tövek sosem változnak, de a kettes típusú toldalékoknak a keményjeles változatát<sup>4</sup> (pl. udm. *canez* + *ĕc*: *canezĕc* 'csizma-k'), a négyes típusú toldalékoknak az *e*-s változatát (pl. udm. *canez* + *ə*: *canezə* 'csizma-m') veszik fel. Ez a B típus.

### 2.3. A lágy affrikátákra végződő tövek

Ezek azért különlegesek, mert lágyságukat nem a lágyságjel (vagy a lágyítást is jelölő magánhangzójel) jelöli. Az udmurtban (*ʉ*, *ʒ*) ugyanúgy viselkedik, mint az előző csoport (pl. *ʒyʉ* + *ĕc*: *ʒyʉĕc* 'orosz-ok'), a komiban (*ʉ*, *ɖʒ*) viszont a kettes típusú toldalékok előtt nem keményjel, hanem lágyságjel jelenik meg (pl. *poʉ* + *ʉc*: *poʉʉc* 'orosz-ok'). Az előzőhöz való nagyfokú hasonlósága miatt B1-es típusnak nevezzük.

### 2.4. A lágy párral rendelkező kemény mássalhangzóra végződő tövek

A *ɖ*, *ʒ*, *l*, *n*, *c*, *m*<sup>5</sup> betűre végződő tövek tartoznak ide. A kettes típusú toldalék keményjeles változata járul hozzá, a hármas típusnak a komi *i*-s, udmurt *ü*-s változata (pl. komi *cem* + *u*: *cemi* 'ad-tam', udm. *cĕm* + *u*: *cĕmü* 'ua. '), a négyes toldaléktípus viszont alapalakjában marad (pl. udm. *cĕm* + *ə*: *cĕmə* 'ad-VxSg3').

<sup>4</sup> Ez csupán az orosz minta mechanikus másolása, hiszen – mivel ezeknek a mássalhangzóknak palatális változatuk nincs – egyértelmű, hogy a *ʉ*, *e*, *ĕ*, *ʉ* utánuk csak a [j]-t jelölheti.

<sup>5</sup> Ismét nem tartoznak ide az affrikáták.

Ez a C típus.

### 2.5. A lágyságjelre végződő tövek

A lágyságjel az egyes és a négyes toldaléktípus előtt kiesik, és a toldalék másodlagos alakja járul a tőhöz (pl. udm. *нинь* + *о*: *нинĕ* 'fog-as', *узъ* + *э*: *узе* 'alsz-ik'). A hármas típus előtt is kiesik a lágyságjel, de ott a toldalék alapalakja kapcsolódik a tőhöz (udm. *мадь* + *уз*: *мадиз* 'mesél-t'). Ez a D tőtípus.

### 2.6. A *Vŭ*-re végződő tövek

Az egyes és a négyes toldaléktípus előtt a *ŭ* kiesik, és a toldalék másodlagos változata jelenik meg (pl. udm. *өӧӱ* + *о*: *өӧӱ* 'vaj-as', *ваӱ* + *э*: *вае* 'hoz-VxSg3'). Ez a tőtípus nagyon hasonlít az előzőhöz, ezért a D1 nevet kapja.

### 2.7. A *Cŭ*-re végződő tövek

Ilyen szótárazott névszói tövek nincsenek, de ige-tövek vannak, ezen kívül a komiban gyakori /j/-betoldásos töveknek /j/-s allomorfiái is különböző írásképpel rendelkeznek. Az ilyen tövekhez sosem járul mássalhangzóval kezdődő toldalék, azaz a kettes típusú sem.

Ha az egyes vagy a négyes toldaléktípus kerül utána, a *ŭ* kiesik, helyére keményjel kerül, és a toldalék másodlagos változata jelenik meg (pl. udm. *бырӱ* + *он*: *бырӱӓн* 'választ-ás'). A hármas típus viszont nem vált ki váltakozást (pl. udm. *бырӱ* + *уз*: *бырӱуз* 'választ-ott'). Ez a típus az E típus.

A komiban elvileg számolnunk kellene azzal az esettel, amikor a *tő* a *чӱ*, *дзӱ* kapcsolatra végződik, és a B1-es típushoz hasonlóan a keményjel helyett lágyságjellel kellene számolnunk. Mivel azonban a komiban a palatális mássalhangzó után álló /j/ a nyelvtörténet során teljesen hasonult az előtte álló hanghoz, ilyen esettel nem találkozunk.

## 3. A kapcsolódási szabályok összefoglalása

A táblázatban az első helyen a *tő*, a második helyen a toldalék alternációját jelöljük: a mínusz az alapalak, a plusz a másodlagos alak megjelenését jelenti, az egy esetben előforduló ++ pedig egy további alakváltozat megjelenését. Láthatjuk, hogy csak egy olyan eset van, amikor egyedül a *tő* alternálódik, a többi esetben vagy csak a toldalék, vagy a toldalék és a *tő* együttes alternációját figyelhetjük meg. A tövek és toldalékok típusokba sorolásával, kapcsolódásaik

szabályainak leírásával a kaotikus kép ugyan nem sokat egyszerűsödik, de kezelhetővé válik.

| tőtípus / toldaléktípus   | <i>a</i> / <i>я</i><br>1. | <i>я</i> / <i>ья</i> / <i>ья</i><br>2. | <i>u</i> / <i>i</i> ~ <i>ü</i><br>3. | <i>э</i> / <i>e</i><br>4. |
|---|---------------------------|--|--------------------------------------|---------------------------|
| A ( <i>V</i> )  | -/-                       | -/-                                    | -/-                                  | -/-                       |
| B ( <i>б</i> , <i>в</i> , <i>г</i> stb.)                              | -/-                       | -/+                                    | -/-                                  | -/+                       |
| B1 (komi <i>ч</i> , <i>дз</i> )                                       | -/-                       | -/++                                   | -/-                                  | -/+                       |
| C ( <i>д</i> , <i>з</i> , <i>л</i> , <i>н</i> , <i>с</i> , <i>м</i> ) | -/-                       | -/+                                    | -/+                                  | -/-                       |
| D ( <i>б</i> /∅)  | +/+                       | -/-                                    | +/-                                  | +/+                       |
| D1 ( <i>Vü</i> / <i>V∅</i> )  | +/+                       | -/-                                    | -/-                                  | +/+                       |
| E ( <i>Cü</i> / <i>Cб</i> )   | +/+                       | -                                      | -/-                                  | +/+                       |