

# Igei wordnet és igei eseményszerkezet ábrázolása

Kuti Judit<sup>1</sup>, Varasdi Károly<sup>1</sup>, Cziczelszki Judit<sup>1</sup>, Gyarmati Ágnes<sup>1</sup>, Nagy Anikó<sup>1</sup>,  
Tóth Marianna<sup>1</sup>, Vajda Péter<sup>1</sup>

<sup>1</sup> Magyar Tudományos Akadémia, Nyelvtudományi Intézet  
1068 Budapest, Benczúr u. 33.  
{kuti,j, varasdi, judit, agnes, nagy, masa, vajda}@nytud.hu

**Kivonat:** A magyar igei wordnet készítése során az igei jelentések viszonyainak ábrázolásához szükséges kibővíteni a Princeton WordNet által főleg a fővekre kidolgozott relációstruktúrát. Az igéket eseményszerűségekként vizsgáltuk, és számba vettük a magyarban az igei jelentések meghatározásához fontos aspektuális információkat. Bemutatjuk Moens és Steedman eseménynukleusz fogalmát, melyet felhasználva aspektuális szempontból csoportosítottuk a magyar igei synsetek egy részét, így pszicholingvisztikailag releváns információkat is ábrázolni tudtunk. Az aspektuális információk jelölésére a wordnet struktúrájába illő új relációkat vettünk fel. Cikkünkben megemlítjük a magyar igei wordnet számítógépes nyelvészeti alkalmazási lehetőségeit is.

## 1 Bevezetés

Tanulmányunkban a magyar wordnet (továbbiakban "HuWN")<sup>1</sup> igei részének készítése során felmerült néhány olyan nyelv- és szófajspecifikus kérdésre, problémára térünk ki, melyekre a munka alapjául szolgáló wordnetek<sup>2</sup> nem ajánlottak kielégítő megoldást, és bemutatjuk, hogy az ezekre adott megoldásaink mily módon viszonyulnak a szabvánnyá vált wordnet alap(ok)hoz. Mivel a wordnet műfaja a főnévi szófaj jellegzetesen hierarchikus viszonyaira épül, a már elkészült wordnetekben egyrészt mennyiségileg a főnévi rész dominál, másrészt annak relációtípusai adtak mintát az igei relációk felvételéhez is. Egy, a főnévi hierarchiát az igei szófajra leképezni próbáló jelentésábrázolás azonban nem bizonyulhat elegendő keretnek az igék által kifejezett jelentések szemantikai viszonyainak meghatározásához, hiszen ehhez elengedhetetlen az igék eseményszerkezetének megvizsgálása. Tanulmányunkban tehát azt mutatjuk be, hogy a magyar igei wordnetben hogyan valósítjuk meg bizonyos, az ige eseményszerkezetéből adódó,

---

<sup>1</sup> Az adatbázis a 2005 tavaszán indult Magyar Ontológia Építése projekt keretén belül készül, mely a Szegedi Tudományegyetem, a MorphoLogic Kft. és a Nyelvtudományi Intézet közös projektuma (GVOP – 2004 – 3.1.1.). A Nyelvtudományi Intézet a magyar wordnet igei részének elkészítését vállalta magára. Ezúton is szeretnénk köszönetet mondani a projekt támogatásáért.

<sup>2</sup> Princeton Wordnet (PWN) [3], BalkaNet [7], EuroWordNet [10]

szemantikai relációkat meghatározó információk tárolását, ábrázolását a wordnet műfaja által lehetővé tett relációs keretek között.

Eddigi tapasztalataink szerint igei jelentések meghatározásához, elkülönítéséhez a magyarban szükséges figyelembe venni aspektuális tulajdonságokat is. Ezért a következő részben néhány, az ige eseményszerkezetével kapcsolatos és aspektuális szempontból releváns megállapítás után felvázoljuk Moens és Steedman *nukleusz* nevű eseménystruktúráját. A továbbiakban megmutatjuk, hogy a nukleusz fogalmának felhasználásával egy olyan eszköztár kerül a kezünkbe, amelynek segítségével

- (a) könnyebben helyezhetők el a magyarban lexikalizált jelentések a wordnet hálójában,
- (b) olyan, pszicholingvisztikailag releváns kapcsolatokat tudunk ábrázolni jelentések között, amelyek eddig hiányoztak a magyar WordNet alapját képező wordnetekből, és
- (c) a HuWN esetleges későbbi számítógépes nyelvészeti felhasználása során is hasznosítható információk kerülnek kódolásra.

## 2 Eseményszerűségek<sup>3</sup> és aspektuális jellemzőik

### 2.1 Logikai következtetések igei jelentések között

Mivel a magyarban bizonyos igeekötők gyakran morfológiailag jelölik az ige által kifejezett eseményszerűség aspektusát, ám a HuWN alapjául szolgáló Princeton WordNet az angol nyelv tipológiai különbözősége miatt nem tartalmaz aspektuális információkat, illetve, mint említettük, az alapvető igei reláció, a hipo-hipernima reláció is főnévi mintára készült, szükséges megvizsgálnunk, hogy a wordnet relációinak alapját képező logikai implikációs viszony milyen módon áll fenn igék között. A wordnet módszertana szerint ugyanis a vizsgált morfémák közötti szemantikai viszonyokat bizonyos logikai következtetések elvégezhetőségén keresztül kell meghatározni. Míg azonban például a főnevek esetében a "*minden X-re igaz, hogy: X (egy) N1, tehát X (egy) N2*" séma igazolásán keresztül az  $N1 < N2$  ( $N1$  hiponimája  $N2$ -nek) összefüggés belátható, az igei morfémák ilyen egyszerű rendezésére nincs mód. A logikai kapcsolatok ugyanis csak teljes (potenciálisan igazságértékkel bíró) kijelentések (illetve az azokat kifejező mondatok) között állapíthatók meg, ám a mondatok logikai szerkezetét éppen az igék alakítják ki módosítóikkal, vonzataikkal együtt.<sup>4</sup> Az ige-vonzat viszony azonban erősen aszimmetrikus: a mondat logikai potenciáljáért az ige a felelős; a vonzat ennek csupán

<sup>3</sup> Az *eseményszerűség* terminust az Emmon Bach által bevezetett *eventuality* terminus megfelelőjeként használjuk.

<sup>4</sup> Amikor igék aspektuális tulajdonságairól beszélünk, valójában igei frázisokat kéne említeniük, hiszen az igék önmagukban alulspecifikáltak erre az információra nézve (ld. pl. [9]).

többé-kevésbé passzív szereplője.<sup>5</sup> Ehhez csatlakozik az a nehézség is, hogy az adott mondatból elvégezhető következtetések nagyban függenek mind az igeidőtől, mind pedig az aspektustól.

Az utóbbi jelenséget például a következő példázza: Míg abból a mondatból, hogy *Mari éppen ment át az utca túloldalára, amikor megpillantotta Bélát* nem következik, hogy Mari át is ment az utca túloldalára (vissza is fordulhatott, hogy Bélát üdvözölje), addig abból a mondatból, hogy *Mari átment az utca túloldalára, amikor megpillantotta Bélát* egyértelműen következik, hogy Mari befejezte a megkezdett cselekvését (sőt, még az a pragmatikai implikátúra is jelen van, hogy Mari pontosan azért hajtotta végre a cselekvést, mert Bélát megpillantotta). A két főmondat pusztán aspektuális értékében<sup>6</sup> különbözik: az első esetben progresszív (imperfektív, folyamatos), a második esetben perfektív (befejezett) szemléletű mondatról állunk szemben, és a mondatok eltérő logikai tulajdonságáért éppen ez a felelős.<sup>7</sup> Nyilvánvaló tehát, hogy az a kérdés, hogy az "át+megy" ige-kötő-ige pár milyen következtetésekben vehet részt, nem választható el attól, hogy milyen aspektuális értékkel szerepel a vizsgált mondatokban. Az (igekötős) ige aspektuális lehetőségei tehát – és vele együtt azon következtetések összessége, amelyben az igével képzett mondat részt vehet – az ige által kifejezett (néhány elemi típusba sorolható) eseménystruktúra függvénye, melyekről alább részletesen is lesz szó. Ezeket az (eseményontológiai) típusokat a nyelvelmélet számára is használható módon először Vendler Zénó [8] határozta meg, majd az ő osztályozását fejlesztette tovább Emmon Bach (ld. [1]), illetve a számítógépes nyelvészet területén Marc Moens és Mark Steedman [6].

## 2.2 A Vendler- és Bach-féle aspektuális osztályok

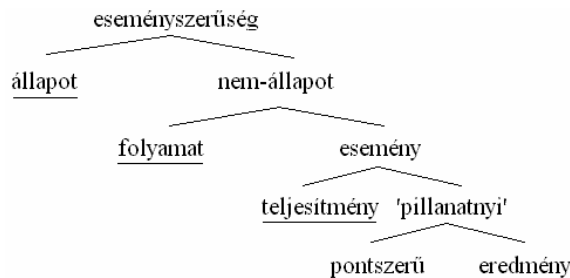
Az eseményszerűségek vendleri felosztása négy aspektuális osztályt különböztet meg arra vonatkozólag, hogy az ige által kifejezett esemény milyen belső időszerkezetet fejez ki. A vendleri osztályozás szerinti négy fő eseménytípus vonzatokkal és kontextussal kiegészülve különböző aspektusokat vehet fel: a folyamatok (pl. úszik) tipikusan a progresszív aspektust, a teljesítmények (pl. kimegy (a szobából)) a progresszív és a perfektív aspektus, az eredmények (pl. kidurran), pedig tipikusan a perfektív aspektust. Az állapotok osztálya sem a progresszív, sem a perfektív aspektust nem tudja felvenni. A Bach által továbbfejlesztett és kiegészített felosztás egy bináris rendszerben elhelyezve az 1. ábrán látható módon jeleníti meg az aspektológiai kategóriákat, kiemelve olyan pontszerű események meglétét, amelyek elkülönülnek az eredményektől (pl. *kattan*):

---

<sup>5</sup> A mondataspektus kialakításában tárgyias ige esetén az ige tárgya is részt vesz. Azonban a tárgy hatása az aspektusra meglehetősen jól megjósolható az ige eseményszerkezetéből és a tárgy sajátosságaiból kiindulva, így azzal a wordnet keretei között külön nem kell foglalkozni.

<sup>6</sup> Az aspektus fogalmát Kiefer (2000)-rel (ld. [4]) összhangban értelmezzük.

<sup>7</sup> A fenti jelenséget a szakirodalom egyébként az "imperfektív paradoxon" néven ismeri, ld. [2].

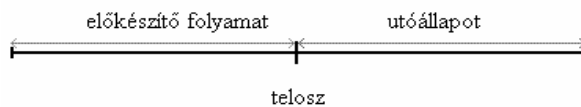


1. **Ábra** Eseményszerűségek Bach-féle aspektuális felosztása

A négy vendleri alosztályt az a tulajdonság is jellemzi, hogy időtartományuk osztható-e avagy sem – azaz, hogy az adott ige által denotált esemenyszerűség az időtartomány legtöbb résztartományára is érvényes-e. Ennek értelmében a *folyamatok* (activity), *állapotok* (state), *teljesítmények* (accomplishment) és *eredmények* (achievement) közül az első kettő homogén esemenyszerűségnek tekinthető, hiszen olyan predikátumok fejezik ki őket, melyeknek tetszőleges részei maguk is ugyanazon predikátum által leírhatók. A *teljesítmények* és *eredmények* ellenben e tekintetben heterogén esemenyszerűségek: különböző jellegű eseménykomponensek koherens egységei. A *pontoszerű* eseményeket szintén nem komplex esemenyszerűségeknek tekintjük. A magyar igei wordnet elkészítésének szempontjából a komplex esemenyszerűségeknek – *eredmények*, *teljesítmények* – az aspektuális jellemzőihez hű ábrázolása érdekes. Ezeknek komplexitását a Moens és Steedman által bevezetett nukleusz struktúra segítségével hívásával értelmezhetjük.

### 2.3 Moens és Steedman eseménynukleusza

Moens és Steedman [6] az esemenyszerűségeknek egy, a vendleri [8] felosztáson alapuló, de annál finomabb rendszerét vezeti be, melynek központi fogalma az *(esemény)-nukleusz*, más néven *triád*. A *triád* név utal arra, hogy egy idealizált esemenyszerűség potenciálisan három összetartozó komponensből áll: *előkészítő folyamat*, *telosz/sikerpont*<sup>8</sup>, és *utóállapot*.



2. **Ábra** A Moens és Steedman-féle eseménynukleusz / triád

Az eseménytriádot rendezett hármasként is lehet ábrázolni:

<a, b, c>

<sup>8</sup> Moens és Steedman a *telosz* kifejezés helyett gyakrabban használja a *kulmináció* (*culmination*) kifejezést. Mi azonban az előbbi terminus mellett maradunk, annak érdekében, hogy a *sikerpontot* megkülönböztessük a Moens és Steedman által szintén kulminációnak nevezett vendleri *eredményektől* (ld. 2. Táblázatot).

ahol a = ELŐKÉSZÍTŐ FOLYAMAT, b = TELOSZ, c = UTÓÁLLAPOT. Moens és Steedman ezt az esemény-alapegységet a lexikalizált igei szófajjal megjelenített verbális nyelvi szint fölött helyezi el. Az esemény-nukleusz egyes komponenseit tehát nem egyes nyelvekben lexikalizált elemek, hanem metanyelvi elemek töltik ki<sup>9</sup>.

metanyelvi szint	<a, b, c>
nyelvi szint	lexikalizált nyelvi elemek, (synsetek)

### 3. Ábra Az eseménynukleusz metanyelvi és nyelvi szinten való megjelenése

A három nukleusz-eseménykomponens<sup>10</sup> egységként való kezelésének létjogosultságára a következő magyarázat adható: Lévéen, hogy aspektológiai szempontból vizsgáljuk az eseményszerűségeket, releváns információnak kell tekintenünk, hogy ha egy-egy tetszőleges lexikalizálódott nyelvi kifejezést aspektológiai jellemzőkre érzékeny nyelvi tesztekkel<sup>11</sup> (a magyarban a progresszív és a perfektív aspektus tesztheivel) vizsgálunk, legfeljebb a fent leírt három komponens együttes megléte mutatható ki. Egy adott eseményszerűséget leíró igeről aspektuális szempontból tehát az mond el valamit, hogy ezen fölöttes esemény-komponensek közül melyeket tölti ki *konceptuálisan*. A „*kimegy a szobából*” igei frázissal megnevezett eseményszerűség példáján bemutatva:

A progresszvizálhatóság az első komponens meglétét teszteli. Egy kifejezés ugyanis akkor és csak akkor elfogadható progresszívben, ha a hozzárendelt triád első komponense konceptuálisan ki van töltve. A triád első komponensének konceptualizáltságát tesztelhetjük a következő mondattal:

*János éppen ment ki az épületből, amikor találkoztam vele.*

A perfektiválhatóság a harmadik komponens meglétét teszteli, amely gyakorlatilag együtt jár a második komponens meglétével (ld. [6]). A magyar nyelv sajátosságai miatt a magyar mondat angolra való fordításával, és Present Perfect igeidőbe való áttételével tesztelhetjük legkönnyebben, hogy egy magyar igei jelentés mögött konceptualizálódott-e a triád második, ill. harmadik komponense.<sup>12</sup>

*By the time Sue arrived, John has gone out of the building.*

A két teszt eredményeképp megállapíthatjuk, hogy a *kimegy az épületből* frázis tehát mindhárom triád-komponenst konceptualizálja:

<sup>9</sup> Mivel ezekre csak nyelvi elemekkel tudunk utalni, jelölési konvencióként a kis kapitálisokkal való írásmódot fogjuk alkalmazni, hogy elkülönüljenek a kurzívval szedett, nyelvi elemektől.

<sup>10</sup> Az eseménykomponenseket most nyelvi lexikalizáltságuktól függetlenül tekintjük.

<sup>11</sup> A perfektivizálás és progresszvizálás lehetséges tesztjeiről ld pl. [4]. Az itt alkalmazott két teszt részben Kiefer (2000)-n, részben saját elgondoláson alapul.

<sup>12</sup> Az angol fordítás ellenére a magyar anyanyelvünk elég erős befolyásoló tényező ahhoz, hogy paradox módon az angol nyelvi mondat helyességét is magyar nyelvi intuíciónk alapján ítéljük meg, s így, bár angol nyelvű tesztel, de a magyar jelentésről nyerjük információt.

*kimegy az épületből* <MEGY A KAPU FELÉ, ÁTLÉP A KAPUN, KINT VAN AZ ÉPÜLETBŐL>

Moens és Steedman a triád komponenseinek a meglétével ill. hiányával pontosítja a vendleri ill. bachi rendszerben már megnevezett kategóriákat. Hogy lássuk hogyan viszonyul a triádok által kialakított osztályozás a Vendler-féle osztályokhoz, figyeljük meg az 1. táblázatot. Ebben a Moens és Steedman rendszerében használt szempontok (atomi ill. kiterjedt és utóállapottal rendelkező, ill. anélküli eseményszerűségek) szerinti csoportosítás látható, mely egyben explicit módon utal a vendleri és Bach-féle aspektuális osztályokkal való megfelelésre (ahol az újabb terminológia eltér a hagyományostól, ott zárójelben megadjuk a régebbit).

1. Táblázat:

Eseményszerűségek felosztása Moens és Steedman rendszerében

	"nem-állapot"		állapot
	atomi	kiterjedt	
+utóállapot	<b>kulmináció</b> (= EREDMÉNY) <i>felismer, megpillant, megnyeri a versenyt</i>	<b>kulminált folyamat</b> (= TELJESÍTMÉNY) <i>épít egy házat, felmegy a hegyre</i>	<i>ismer, szeret, tud, hasonlít</i>
- utóállapot	<b>pontszerű esemény</b> <i>vakkant, dörren, koppan, megbotlik</i>	<b>folyamat</b> <i>fut, úszik, sétál, zongorázik</i>	

A nukleusz eseménykomponenseinek összetartozása több, mint pusztán időbeli egymásutániság; az okozáshoz, és a lehetővé tételhez hasonlatos, de azokkal nem megegyező viszony, melyet Moens és Steedman *contingency*-nek nevez. A három nukleusz-komponens kölcsönös dependenciája miatt egyiket sem lehet saját jogán előkészítő *folyamatnak*, *sikerpontnak* illetve *utóállapotnak* tekinteni. Az olyan eseményszerűség, amely a fenti tesztek alapján látszólag rendelkezik *előkészítő folyamattal*, ám nincs sem *sikerpontja*, sem *utóállapota*, (jelölése lehetne  $\langle a, \emptyset, \emptyset \rangle$ ) nem tekinthető *előkészítő* folyamatnak, hiszen nem készít elő semmit. Az ilyen eseményszerűség a vendleri *folyamat* kategóriának feleltethető meg. Az olyan eseményszerűség, amely a fenti tesztek alapján látszólag rendelkezik *utóállapottal*, ám nem rendelkezik *sikerponttal* (jelölése lehetne  $\langle \emptyset, \emptyset, c \rangle$ ), értelemszerűen nem lehet *utóállapot*, hanem a vendleri értelemben vett *állapotnak* tekintendő. Az olyan eseményszerűség pedig, mely *előkészítő folyamat* és *utóállapot* nélküli *sikerpontnak* tűnik, szintén nem nevezhető annak, hanem a bachi értelemben vett *pontszerű eseménynek* kell tekintenünk (jelölése lehetne  $\langle \emptyset, b, \emptyset \rangle$ ). Összefoglalva tehát elmondhatjuk, hogy a triád kitöltött *utóállapota* implikálja egy *sikerpont* meglétét, a kitöltött *sikerpont* pedig implikálja egy *előkészítő folyamat* meglétét.

A nukleusz egyes komponenseinek konceptuális kitöltöttsége szerint tehát potenciálisan  $2^3$  számú különböző aspektuális kategória különböztethető meg. Ezekből az imént említett hármat (folyamat, pontszerű esemény, állapot), mint nem komplex eseményszerűséget, Moens és Steedman nem tárgyalja részletesen, mi azonban foglalkozunk velük a HuWN-ben. A lehetséges kombinációs lehetőségek

közül további hármat világismereti okokból kizárhatunk: A mindhárom komponens kitöltetlenségét jelentő kombináció sem nyelvileg, sem konceptuálisan nem lehet releváns számunkra; a kizárólag előkészítő folyamattal és telosszal rendelkező eseményszerűség (jelölése lehetne <a, b, Ø>), csakúgy mint az *előkészítő folyamattal* és *utóállapottal* rendelkező eseményszerűség (jelölése lehetne <a, Ø, c>), a *telosz* és az *utóállapot* összetartozása miatt nem lexikalizálódhat.

### 3 Az eseménynukleusz fogalmának felhasználása a HuWN-ben

Moens és Steedman nukleusz-strukturájának segítségével tehát nyelvi lexikalizált kifejezések mögött lévő metanyelvi elemek konceptuális meglétét illetve hiányát állapíthatjuk meg. Az az információ, hogy egy ige egy idealizált komplex eseménységhez képest hány elemet konceptualizál, nem más, mint az eseményszerűség *telikusságára*, ill. *atelikusságára* vonatkozó információ. Amennyiben egy ige által leírt esemény-nukleusz harmadik komponense ki van töltve<sup>13</sup>, *telikus* eseményszerűségről, amennyiben nincs, *atelikus* eseményszerűségről beszélünk.

#### 3.1 A telikusság jelölése a HuWN-ben

Ha a telikusság szempontjából próbáljuk meg áttekinteni a nukleusz egyes komponenseinek kitöltöttsége szerinti öt, nyelvileg is realizált lehetséges mintázatot, a 2.3-ban bevezetett rendezett hármásokat idéző ábrázolásmód tűnik alkalmasnak:

2. Táblázat:

Komplex eseményszerűségek telikussága Moens és Steedman nukleusz-strukturájának segítségével ábrázolva

<i>a triád komponenseinek kitöltöttsége</i>	<i>a triádot lexikalizáló frázis konceptualizált komponenseinek metanyelvi megnevezése</i>	<i>az igei frázis telikussága</i>
<a, b, c>	<i>kimegy az épületből:</i> <MEGY A KAPU FELÉ, ÁTLÉP A KAPUN, KINT VAN>	+ utóállapot → telikus
<Ø, b, c>	<i>kidurran:</i> <Ø, A KIDURRANÁS PILLANATA, KIDURRANT ÁLLAPOT>	+ utóállapot → telikus

Az egyszerű eseményszerűségek (*folyamatok, állapotok, pontszerű események*) közül a *folyamatokat* és az *állapotokat* atelikusnak szokás tekinteni, a *pontszerű események* pedig kontextus nélkül, önmagukban alulspecifikáltak erre az információra nézve.

A HuWN készítésekor arra az igei jelentések rendszerezése során visszatérő kérdésre, hogy kódoljuk-e, és ha igen, milyen formában, a csupán aspektuálisan eltérő, a wordnet szinonima fogalma alapján azonban szinonimnak értékelendő jelentéseket, az eseménynukleusz fogalmának felhasználásával igennel válaszolhatunk. Amennyiben egy-egy, a wordnetben synsetként szereplő jelentést

<sup>13</sup> Amint az előző pontban említettük, a harmadik eseménykomponens kitöltöttsége komplex eseményeknél magával vonzza a második komponens kitöltöttségét is.

minimális propozícióvá alakítunk, meg tudjuk mondani róla, hogy az általa lexikalizált nukleusz utóállapota ki van-e töltve, vagy sem.<sup>14</sup> Azáltal, hogy egy jelentésről kódoljuk, hogy van-e utóállapota (és ezáltal telosza is) vagy sem, és hozzárendelünk egyet a triád-komponensek konceptualizáltságának megfelelő öt minta közül, egyértelműen leolvasható lesz az ige által kifejezett eseményszerűség telikus ill. atelikus volta. Ezt az információt a wordnetben a synsetekben tároljuk, oly módon, ahogyan maga a vonzatkeret-információ is tárolásra kerül: megjelöljük, hogy a triád három komponense közül melyek konceptualizálódtak a magyarban.

A három, a 2.2 és 2.3. alatt említett egyszerű eseményszerűség esetében is, ahogy fentebb utaltunk rá, az átláthatóság és egységesség kedvéért megtartjuk az aspektuális információ rendezett hármasként való jelölési konvencióját. Ennek megfelelően pl. a {fut} synset esetében a rendezett hármasként való jelölésből (<a,  $\sqcap$ ,  $\sqsupset$ >), ahol  $\sqcap$  a nem konceptualizálódott triád-elemekre utal, leolvasható az az információ, hogy az eseményszerűség atelikus, de leolvasható az is, hogy az eseményszerűség egy vendleri folyamat, amire az önmagában betöltött előkészítő folyamat pozíció utal.

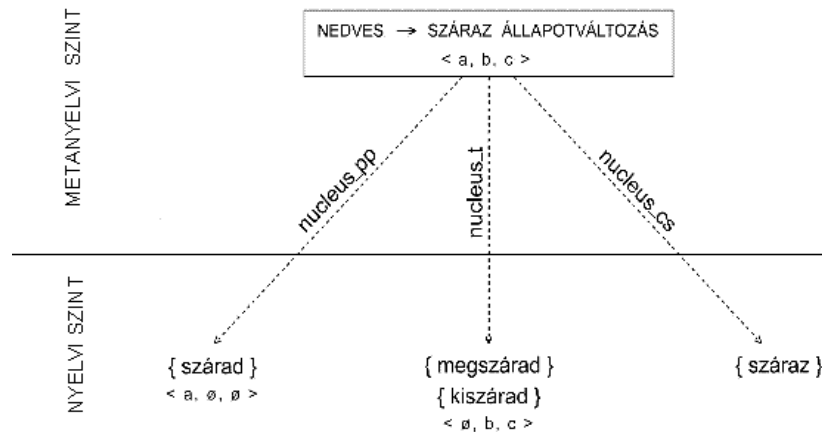
### 3.2 Komplex eseményszerűségek az igei HuWN-ben

Azon túl, hogy egyes igei synsetekben az adott jelentésre vonatkozó minimális aspektuális információt tárolunk, a wordnet relációs struktúrája és az eseménynukleusz egységként való felfogása megengedi, hogy olyan komplex eseményszerűségek esetében, amelyeknek bizonyos triád-komponensei nemcsak koncepuálisan vannak kitöltve, de nyelvileg is lexikalizálódtak, megjelenítsük az egyes eseménykomponensek összetartozására vonatkozó információt is. Ennek megfelelően annak analógiájára, ahogyan PWN mellékeveinek a főnévi hierarchiától független strukturálása vált elfogadottá, az igei wordnet készítésekor is példaként lehet venni a melléknévi synsetek jelentős részének szófajspecifikus (bokros / clusteres) elrendezését. Az imént felvázolt triád ugyanis leképezhető a wordnet rendszerére relációs formában. Ez annyit jelent, hogy a Moens és Steedman-féle eseménytriád által leírt metanyelvi szintnek, ahol tudjuk, megfeleltethetjük a nyelvi szinten lexikalizálódott elemeket – amelyeket a wordnetben synsetek jelölnek. A két szint összekapcsolását a 4. ábra mutatja.

---

<sup>14</sup> A wordnet synseteiben előforduló jelentések minimális propozíciókká alakíthatóságát az biztosítja, hogy egy-egy igehez hozzárendeljük az adott jelentésben előforduló összes vonzatkeretét. Sokszor egy-egy vonzatkeret-bejegyzés több összevont vonzatkeretet takar, opcionális argumentumokkal. Ilyenkor az igeiket a minimálisan kötelező argumentumszámukkal értelmezzük. Pl. az opcionális tárgyval felvett *eszik* vonzatkeret esetében a tárgy nélkül képzett minimális predikátumról mondhatjuk ki atelikus voltát.





4. Ábra Moens és Steedman eseménynukleuszának leképezése a wordnet synseteire

A HuWN-ben a synsetek strukturálhatóságának céljából bevezetett mesterséges csomópontok (ld. [5]) alkalmasak arra, hogy egy-egy, a metanyelvi szinten elhelyezett eseménynukleuszt megnevezzünk általuk – a fenti példában a nedves és száraz állapotváltozást denotáló komplex eseményszerűséget.<sup>15</sup> A wordnetek relációs felépítése lehetővé teszi, hogy annak megfelelően, hogy egy lexikalizálódott igei kifejezés a triád mely komponensét lexikalizálja, kapcsoljuk külön-külön névvel a triád egységét megjelenítő mesterséges csomópontokhoz. A triád komponenseinek eredeti, angol nevét alapul véve a három felvehető reláció neve *nucleus\_pp* (a *preparatory process* kifejezésből adódóan), *nucleus\_t* (a *telosz* kifejezésből adódóan), és *nucleus\_cs* (a *consequent state* kifejezésből adódóan). A három reláció közül a *nucleus\_pp* értelemszerűen az előkészítő folyamatot lexikalizáló synset felé mutat, a *nucleus\_t* a sikerpontot lexikalizáló synset felé, míg a *nucleus\_cs* az utóállapotot lexikalizáló synset felé.

Ezáltal olyan, az angolban egyetlen igeként lexikalizálódott jelentések különíthetők el, amelyek a magyarban gyakran egy igekötős, aspektuális többletinformációt hordozó, és egy igekötő nélküli, aspektuálisan aluspecifikáltabb igeként jelennek meg. A fenti példában mind a {szárad}, mind a {megszárad} synset az angol {dry:2} synsetnek feleltethető meg. A nukleusz struktúra wordnetbe való leképezése nélkül a {megszárad} synset az eddigi, rendelkezésünkre álló relációk felhasználásával kizárólag a {szárad} synset hiponimájaként lenne felvehető. Ez a tárolási mód azonban összemosná az említett két jelentés között fennálló implikációs viszonyt az argumentumokra vonatkozó megszorítás különbségeire építő hiponima-hipernima viszonytal (amely pl. a {hervad, fomnyad} és a {rohad} synset között áll fenn, mivel előbbinek csak növényi alanya lehet, utóbbinak pedig nemcsak az). A nukleusz struktúra wordnetbe való leképezésével a két igei synset között nem szükséges explicit módon egy külön relációt felvennünk: az őket összefogó mesterséges csomóponton keresztül, a *nucleus\_pp* és a *nucleus\_t* relációk mentén egyértelműen

<sup>15</sup> A mesterséges csomópontokat a wordnetben a kapitálisokkal való írásmóddal különböztetjük meg a természetes nyelvi synsetektől.

meghatározható, hogy a {szárad} synset a {megszárad} synset előkészítő folyamata, így per definitionem a {megszárad} implikálja a {szárad}-ot, de nem fordítva.<sup>16</sup> Az egy triádon belül lexikalizált igék esetében tehát az igekötős és igekötő nélküli alak elhelyezése a wordnetben sok esetben kézenfekvőbb, mint az eddig rendelkezésünkre álló relációk segítségével. Ezenkívül a fenti példában a triád harmadik komponensét lexikalizáló melléknévi synsethez ({száraz}) mutató *nucleus\_cs* relációnak köszönhetően egy olyan, pszicholingvisztikailag releváns információ lesz leolvasható a {szárad}, {megszárad}, ill. a {száraz} synsetek között, mely az angol wordnetet alapul véve, a triád beépítése nélkül elveszne.

### 3.2.1 Jellegzetes triádok az igei HuWN-ben

Lévén, hogy a Moens és Steedman féle eseménynukleusz mindhárom komponensének kitöltött állapota egy *előkészítő folyamattal*, egy *sikerponttal* és egy *utóállapottal* rendelkező eseményszerűséget feltételez, abból a feltevésből indultunk ki, hogy a triád-struktúra adaptálása a wordnetbe a valamilyen változást (leginkább talán állapotváltozást) jelentő igék körében nagy százalékkal bizonyul majd hasznosnak. Ennek megfelelően a {változik:1} ill. {változtat:1} csomópontokat – melyek a PWN csomópontjainak önmagukban is egy negyedét kiteszik – választottuk ki arra a célra, hogy megvizsgáljuk a triád-struktúra gyakorlati beépítésének előnyeit a HuWN-be.

Az a sejtésünk, hogy a {változik:1} ill. {változtat:1} synsetek hiponimáiból álló részfákban számos olyan csomópontot találhatunk, amelyek alkalmasak a triádban való ábrázolásra, beigazolódott, amikor egy-egy esemény-nukleusz komponenseinek kitöltöttségét kódoltuk. Eredményeinket a 3. táblázatban bemutatott adatok illusztrálják.

3. Táblázat:

A felvett nukleuszok aránya a {változik:1} ill. {változtat:1} csomópontok alatt

nukleusz-típus	{változik}	{változtat}
<*, b, c>	24%	14%
<a, b, c>	30%	42%
Összesen	54%	56%

A {változik} ill. {változtat} synsetek közvetlen első ill. második szintű hiponimái közül összesen 150-et megvizsgálva tehát arra jutottunk, hogy triádok felvétele az esetek több mint felében tette könnyebbé a magyarban lexikalizálódott jelentések ábrázolását a wordnetben.

A {változik} és {változtat} csomópontok (ill. megfelelő hiponimái) közötti konceptuális hasonlóság miatt felmerült, hogy az ezekhez tartozó eseménynukleuszok párhuzamos módon épülnek fel. Közelebbről megvizsgálva azonban a triádok egyes pontjai közti viszonyok különböznek, amit a kauzatív-inchoatív alternációval magyarázhatunk. A különbség a következő: a <SZÁRAD, MEGSZÁRAD, SZÁRAZ> triád második és harmadik eleme között *következmény*-reláció áll fenn, amennyiben a

<sup>16</sup> Ld. a 2.3 alatti leírást a Moens és Steedman által *contingency*-nek nevezett összetartozásról a triád elemei között.

*megszárad* igével kifejezett esemény következménye a *megszárad* alanyának száraz volta. A látszólag párhuzamosan felépülő <SZÁRÍT, MEGSZÁRÍT, TÚLVAN A MEGSZÁRÍTÁS FOLYAMATÁN> triád második és harmadik komponense között más a viszony: az eseményszerűség utóállapota a *megszárít* ige tárgyára vonatkozik, azaz a {megszárít:1} synset okozás (*causes*) relációban áll a {száraz} melléknévi synsettel.

Bár az angol oldalon a {dry:1} (kauzatív) és {dry:2} (inchoatív) synsetek között fel van véve a *causes*-reláció<sup>17</sup>, sem a {dry:1}, sem a {dry:2} synset felől nem mutat semmilyen reláció a melléknévi {dry:1} felé. Ha a magyar wordnetben a két triád viszonyának jelölése az angol mintára történne, mind az előkészítő folyamatot lexikalizáló kauzatív / inchoatív igepár ({szárít} - {szárad}), mind a sikerpontot lexikalizáló kauzatív/ inchoatív igepár ({megszárít} - {megszárad}) között fel kellene venni egy-egy *causes*-relációt, ami azonban szükségtelenül megduplázná a felveendő relációk számát, s a kauzatív igék és a melléknévi synsettel lexikalizált okozat között még mindig csak közvetetten lenne a kapcsolat leolvasható. Ezért az egyes triádokat jelölő mesterséges csomópontokat kötjük össze *causes*-relációval, azaz: MEGSZÁRÍT -- *causes*--> MEGSZÁRAD, aminek köszönhetően éppúgy kiszámítható a triád élei mentén a kauzatív triád és az inchoatív triád igéi között fennálló okozás-reláció, mint a kauzatív triád igéi és az inchoatív triád utóállapotát lexikalizáló melléknévi synset között fennálló okozás-reláció.

## 4 Alkalmazási lehetőségek

Amellett, hogy a nyelvi elemek idioszinkratikus tulajdonságainak egységes ábrázolása jól beleillik a wordnetek fő feladatai közé, a HuWN ilyen típusú bővítése gyakorlati haszonnal is jár. Mint láttuk, az igék telikus-atelikus ill. befejezett-folyamatos volta könnyen meghatározható a triád komponenseinek segítségével. A HuWN ezen információinak felhasználása javíthat egy (magyar-angol) gépi fordító működésein, például az igeidők pontosabb megfeleltetése terén. Különösen hasznos lehet ez, ha a magyarban morfológiailag meglévő két igeidőt vetjük össze az angol igeidőkkel. A magyarra jellemző, hogy a morfológiailag jelen időben megjelenő telikus igealakok jövő idejű referenciával bírnak. A *Felhívom Pétert* mondatnak angolul nem *I call Peter*, hanem *I will call Peter* lenne a helyes fordítása. Hasonlóan, a folyamatos aspektussal bíró múlt idejű alakok nem egyszerű, hanem folyamatos múlttal fordítandók angolra: *Péter az udvaron játszott* megfelelője *Peter played in the court* helyett *Peter was playing in the court*. Az aspektuális információk mondatgenerálásnál is felhasználhatók – legyen szó akár angol-magyar fordításról, akár más, generálást igénylő feladatról.

Az igék ezen tulajdonságai a gépi mondatmegértésben is segíthetnek. Egy számítógép belső reprezentációjának fontos eleme az igék ezen idioszinkratikus tulajdonságainak ismerete. Enélkül, pusztán az esetleg a mondatban meglévő időhatározókat figyelembe véve nem lehet pontos egy történet időbeli struktúrájának ábrázolása.

---

<sup>17</sup> Sajnos a {change:2} és {change:1} csomópontok közötti ilyen kapcsolatmegadás nem konzekvens a PWN-ben.

## **Bibliográfia**

1. Bach, E.: The Algebra of Events. *Linguistics and Philosophy* 9 (1986) 5-16
2. Dowty, D.: *Word Meaning and Montague Grammar*. D. Reidel, Dordrecht, W. Germany (1979)
3. Fellbaum, C.: *WordNet An Electronic Lexical Database*. MIT Press (1998)
4. Kiefer, F.: *Jelentélmélet*. Corvina, Budapest (2000)
5. Kuti, J., Vajda, P., Varasdi K. Javaslat a magyar igei WordNet kialakítására. In III. Magyar Számítógépes Nyelvészeti Konferencia - MSZNY 2005, Szeged, 2005, 79-88.
6. Moens, M., Steedman, M.: Temporal ontology and temporal reference. *Computational Linguistics* Volume 14 Issue 2 (1988) 15-28
7. Tufis, D. et al. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. In: *Romanian Journal of Information Science & Technology*, 7(1-2) (2004) 1-35.
8. Vendler, Z.: Verbs and Times. *Philosophical Review* 66 (1957) 143-160
9. Verkuyl, H. J.: *On the compositional nature of the aspects: Foundations of Language*. Supplementary Series 15. Reidel, Dordrecht (1972)
10. Vossen, P.: *EuroWordNet General Document*. Technical Report EuroWordNet (LE2-4003, LE4-8328) (2005)