

Hungarian WordNet and representation of verbal event structure*

Judit Kuti[†], Károly Varasdi[†], Ágnes Gyarmati[†] and Péter Vajda[†]

Abstract

During the development of the Hungarian WordNet we found that the primarily hierarchical structure of the Princeton WordNet (PWN) had to be expanded in order that it could be used for a more accurate representation of relations among verbal meanings in Hungarian, as well. Treating verbs as eventualities we made some aspectual information explicitly available in the representation of verbs. Using the notion of *nucleus* introduced by Moens&Steedman we added new relations to the WordNet and classified verbal synsets according to aspectual characteristics. This enabled the representation of some psycholinguistically relevant pieces of information and a wider possible usage of the thus extended Hungarian verbal WordNet in the field of computational linguistics.

Keywords: WordNet, verb, event structure, event ontology, aspect

1 Introduction

In the present study we examine some specific problems related to Hungarian verbs which we have encountered when developing the Hungarian WordNet (HuWN), and show how our results relate to the Princeton WordNet, the standard database we relied on when building the HuWN. As wordnets¹ were originally designed to describe the hierarchical structure of nouns, it is nouns that constitute a preponderant

*The building of the Hungarian WordNet was carried out in the framework of the project "Building of the Hungarian WordNet" (GVOP – 2004 – 3.1.1.) since 2005. The project was a collaboration between the University of Szeged, MorphoLogic Ltd. and the Research Institute for Linguistics of the Hungarian Academy of Sciences. The Research Institute for Linguistics was involved in building the verbal part of the WordNet. It is here we would like to thank for the support of the project. As the present article largely relies on the common work carried out in cooperation with our colleagues Judit Cziczelszki, Anikó Nagy and Marianna Tóth, we would like to thank them for their help and contribution.

[†]Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, E-mail: {kuti,j,varasdi,aagnes,vajda}@nytud.hu

¹When talking about a specific WordNet of a given language, we refer to it with the widespread, trademark-like spelling, using capital 'W' and 'N', while when referring to the database type as to a common noun we use minuscules.

part of existing wordnets and it is nominal relations that, to a large extent, served as examples for verbal relations, as well.² However, the choice of the two distinct names for equivalent relation types in the two respective parts of speech in PWN indicates already that a meaning representation framework for verbs cannot be solely designed on the basis of the existing grounds for a nominal hierarchy, not even in the case of a language like English, in which verbs as lexical units bear little or no information related to aspect or aktionsart.³ In the case of languages in which this information is stored in the verb, e.g. through preverbs, the inapplicability of the nominal structure to the verbal network is even more obvious, when attempting to develop a lexical semantic network.

Examining the event structure of verbs provides help in approaching these questions. Accordingly, in what follows, we would like to show what ways of representing certain pieces of information that stem from the event structure of verbs and determine their semantic relations we have worked out – within the framework facilitated by wordnet as a genre.⁴ We present some fundamental statements on event structure and aspectuality in general, on aspect and aktionsart of verbs with special respect to Hungarian, and an elementary event-structure called *nucleus* introduced by Moens&Steedman. The remaining parts of our study show that by using the notion of *nucleus* we acquire a means that enables us to

1. incorporate lexicalised meanings into WordNet more easily than was possible previously
2. represent psycholinguistically relevant information that were so far missing from the Hungarian WordNet
3. store information that prove to be useful for computational linguistic applications of the HuWN.

2 Eventualities and their aspectual properties

2.1 Logical implication between verbal meanings

It is necessary to examine in what way the relation of logical implication holds between verbs⁵ since this is what both the relations *troponymy* and *hyponymy* are based on. Implications of a sentence are highly dependent on its aspect, illustrated by the following examples:

²Although the main relation used in the structuring of verbal synsets in PWN, *troponymy*, as introduced by Fellbaum (see [3]), is, in principle, different from the hypo-hypernymy relation used in the nominal part of the WordNet, the hierarchical relational structure adopted from the nominal part suggests nonetheless enough similarity between the two relations to call the equivalent verbal relation expressing super- and subordination in BalkaNet (see [8]) *hypo-hypernymy*.

³In the definition of these two terms we largely rely on [5].

⁴The slightly theoretical linguistically oriented nature of the paper is intended to be balanced by the usefulness of WordNet for computational linguistic applications.

⁵When talking about logical implications and aspectual properties of verbs, we should, in fact, be talking about verbal phrases, since verbs on their own are underspecified with respect to this kind of information, see [10].

1. *Mari éppen ment át az utca túloldalára, amikor megpillantotta Jánost.*
'Mary was crossing the street when she saw John.'
2. *Mari átment az utca túloldalára, amikor megpillantotta Jánost.*
'Mary crossed the street when she saw John.'

While sentence (1) does not imply that Mary actually crossed the street – she might have turned back to greet John, sentence (2) does imply that Mary did finish crossing the street (moreover, the pragmatic implicature suggesting that Mary crossed the street because she had noticed John, is also present).

The difference between the two main clauses in Hungarian is merely aspectual: the first is in progressive, while the second is in perfective aspect, each possessing a different logical potential.⁶ It is, thus, obvious that the question concerning what implications the preverb and verb as a whole can take part in is not separable from its aspectual value in the sentence. Although in Hungarian the actual aspect of a sentence⁷ is of course determined by many factors in the sentence besides the verb, its aspectual *potential* — as well as the sentences it can imply — is largely determined by the event structure of the verb.

In Hungarian some preverbs can bear information related to both aspect and aktionsart. This alone might make Hungarian seem to be similar to Slavic languages. However, on the one hand, Hungarian does not express aspect in as a predictable manner, as e.g. Russian, whose WordNet we could have used as a basis for the Hungarian one, if the two languages had had enough similarities. On the other hand, aspect and aktionsart in Hungarian are interwoven in a way that is unique among the languages that so far have been developed a wordnet for. The perfective aspect for example goes almost always hand in hand with one of the aktionsart-types that are present in Hungarian (Kiefer 2006:45, see [5]).

Furthermore, Hungarian has an extremely rich system of preverbs which can modify the meaning of the stem, making it inevitable, when dealing with Hungarian, to consider aspectual characteristics as much as possible within the given framework. As already mentioned, the basic verbal relation, hypo- and hypernymy, but troponymy just as well, were elaborated based on the pattern of nominal relations in the sense that the wordnet-methodology requires that semantic relations between morphemes hold through logical implications. While in the case of nouns one can show that N1 is a hyponym of N2 — by checking whether the pattern "it is true for each X that if X is an N1, then X is an N2" holds —, this is not possible for verbs, since one can only establish logical relations between propositions or the sentences expressing them, but the logical structure of sentences is determined by verbs together with their modifiers and complements. However, the verb-complement relation is highly asymmetrical: the logical potential of the sentence is determined by the verb; complements are only more or less passive participants.⁸ As the PWN, which has served as a basis for HuWN, does not contain aspectual information

⁶The above phenomenon is known as the *imperfective paradox*.

⁷Aspect itself is considered a sentence semantic category (see [5]).

⁸In the case of verbs with direct object complements it is also the direct object that takes part in determining the aspect of the sentence. However, the impact of the direct object on the aspect

due to the lack of morphological marking of aspect in English, another way had to be found for representing typically occurring phenomena related to aspect in Hungarian.

A first framework of approaching aspect in general is provided by Zeno Vendler [9], who developed a type of event ontology in a way that would become useful for linguistic theory. This system was later elaborated by Emmon Bach [1], and worked out for computational linguistics by Marc Moens and Mark Steedman [7]. Drawing on Moens&Steedman’s work we would like to suggest a way to structure aspectually related verb meanings in WordNet.

2.2 Aspectual classes according to Vendler and Bach

Vendler’s classification of eventualities distinguishes between four aspectual classes according to the internal temporal structure of the event expressed by the verb. According to Vendler the four event types — with arguments and with context — differ in the aspects they may take: activities (e.g. *swim*) typically take the progressive aspect, accomplishments (*go out of the room*) take both the progressive and the perfective aspect, and achievements (*blow up*) take the perfective aspect. States take neither the progressive nor the perfective aspect. The classification as further developed and extended by Bach represents aspectual categories in a binary system, highlighting the existence of point expressions that are different from achievements (e.g. *click*). In Bach’s terminology Vendler’s accomplishments are called *protracted events*, achievements are called *culminations*, while *point* expressions are called *happenings*.

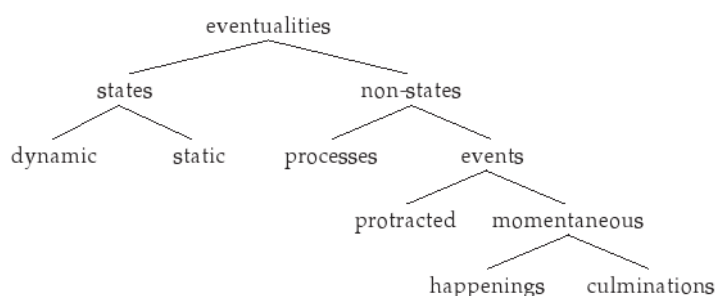


Figure 1: Classification of eventualities according to Bach

Vendler’s four aspectual classes are also characterised by whether the interval of the event is divisible or not — i.e. whether the eventuality denoted by the verb holds for most of the sub-intervals, as well. Accordingly, *activities* and *states* may be considered homogeneous eventualities,⁹ since they are expressed by predicates

can be relatively well predicted from the event structure of the verb and the properties of the object, so we do not have to specifically deal with this in the framework of the WordNet.

⁹We are using the term *eventuality* after Bach, see [1].

any sub-intervals of which may be described by the very same predicates. *Accomplishments* and *achievements* on the other hand are in this respect coherent units of different kinds of heterogeneous event-components. *Point* expressions are also taken to be non-complex eventualities. From the point of view of constructing the Hungarian verbal WordNet it is the representation of complex eventualities — *achievements* and *accomplishments* — in a way that does justice to their aspectual properties that is of interest to us. Their complexity might be interpreted with the help of the so-called *nucleus-structure* introduced by Moens&Steedman.

2.3 The event-nucleus of Moens&Steedman

Moens&Steedman introduce a classification of eventualities relying on Vendler's aspectual classes but further refining it. Their central notion is that of an event-nucleus, which might be called a tripartite structure or triad, as well. The reason for the latter name is that an idealised eventuality consists of potentially three components belonging together: *preparatory phase*, *telos/ culmination* and *consequent state*.

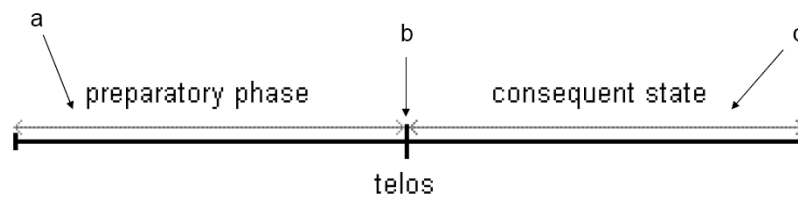


Figure 2: The event-nucleus of Moens&Steedman

One may also represent the triad as an ordered triple $\langle a, b, c \rangle$ where a =preparatory phase, b =telos and c =consequent state. Moens&Steedman place this idealised event-unit beyond the level of linguistically manifested lexicalised meanings. The components of the event-nucleus are thus filled with meta-linguistic and not with lexicalised linguistic elements.¹⁰

meta-language level	$\langle a, b, c \rangle$
linguistic level	lexicalized linguistic units (verbs)

Figure 3: The event-nucleus on meta-language and linguistic levels

Treating the three nucleus-components¹¹ as a unit might be justified as follows. Since we are examining eventualities from an aspectual point of view, the fact that when testing a lexicalised expression with linguistic tests sensitive to aspectual

¹⁰Since we may only refer to these with linguistic elements, we will use small capitals so that they can be held apart from italicised, linguistic elements.

¹¹Here we are dealing with the event-components irrespective of whether they are lexicalised or not.

properties (in Hungarian the tests of the progressive and the perfective) the co-occurrence of no more than the three components outlined above may be shown, must be considered relevant. We may, thus, acquire information about the aspectual properties of a verb expressing a certain eventuality by looking at which of the three event-components described above are conceptually present.

On the example of the eventuality lexicalised with the verbal phrase *go out of the room*: The existence of the first component can be tested by looking at whether the expression can be put into the progressive. An expression will be acceptable in the progressive if and only if the first component of its triad is conceptually present. The existence of the third component, which practically goes hand in hand with the presence of the second one,¹² can be tested by looking at whether the expression can be put into the perfective (see [7]). Due to certain characteristics of the Hungarian language the easiest way we can test whether certain components of the triad are conceptualised is by translating the Hungarian sentence into English and putting the translated equivalent into Present Perfect / Progressive.¹³

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

'János was going out of the building when I met him.'

4. *Mire Zsuzsa megérkezett, addigra János kiment az épületből.*

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

As a result of the two tests we can see that the phrase *go out of the building* conceptualises all the three components of the triad:

<GOES TOWARD THE GATE, PASSES THE THRESHOLD, IS OUTSIDE>

Moens&Steedman elaborate the categories established by Vendler/Bach, by adding the factors of the presence or lack of the triad-components. In order to see how the classification according to the triad-components relates to the classification of Vendler/Bach, let us look at Table 1. This shows the classification of eventualities according to the factors taken into consideration by Moens&Steedman (+/- atomic and +/- consequent state), explicitly referring to the equivalents in Vendler's and Bach's system, where possible (in cases where the new terminology differs from the former one, we have indicated the latter in brackets).

¹²Although the second component may sometimes be called culmination point, this is not supposed to imply that this event-component necessarily has to take place within a pointlike short time. This can indeed be a longer period which is conceptualised as a point.

¹³Since this methodology may be surprising at first, some explanation is in order. In Hungarian — as opposed to English — there are no clear-cut and simple tests that are sensitive enough to the aspectual properties of a sentence (or verb phrase). Realizing that the impossibility of providing a usable test battery for Hungarian, we chose a detour, as it were, through a proxy in English. Benefiting from the situation that everybody in the WordNet developers' team spoke English on an advanced level and had learnt to be sensitive to certain aspectual features in English, we decided to rely on our tacit knowledge of the aspectual features we wanted to test. When translating a Hungarian sentence into the English Present Perfect or Progressive, one had to judge its aspectual acceptability irrespective of whether the translation was correct in any other respect. Obviously, this methodological shortcut should be backed by further research in second language acquisition to be of sound theoretical value, but we believe that used with sufficient care it provides a reliable tool when the tests in the object language prove too complicated for practical usage.

Table 1: Eventualities in the system of Moens&Steedman

	non-states		states
	atomic	extended	
+conseq	culmination (=ACHIEVEMENT) recognize, spot win the race	culminated process (=ACCOMPLISHMENT) build a house eat a sandwich	resemble understand love know
-conseq	point hiccup tap, wink	process run, swim, walk play the piano	

Theoretically 2^3 different potential aspectual types may be distinguished according to the conceptual presence of the nucleus-components, listed as follows.¹⁴

$$\begin{array}{lll}
 \langle \emptyset, \emptyset, \emptyset \rangle & \langle \emptyset, b, c \rangle & \langle a, \emptyset, \emptyset \rangle \\
 \langle a, b, c \rangle & \langle a, \emptyset, c \rangle & \langle \emptyset, b, \emptyset \rangle \\
 & \langle a, b, \emptyset \rangle & \langle \emptyset, \emptyset, c \rangle
 \end{array}$$

The coherence of the nucleus components is more than mere temporal sequentiality, it is what Moens&Steedman call *contingency* — ”a term related, but not identical to a notion like causality” [7]. The mutual dependency among the three components of the nucleus means that none of them can be seen as *preparatory phase*, *culmination* or *consequent state* per se. An eventuality that, based on the above tests, seems to possess a preparatory phase, but lacks both culmination and consequent state (could be marked as $\langle a, \emptyset, \emptyset \rangle$) cannot be seen as a preparatory process, as it does not precede anything. By analogy, an eventuality that, based on the above tests, seems to possess a consequent state but lacks a *culmination* (could be marked as $\langle \emptyset, \emptyset, c \rangle$) cannot be seen as a consequent state, just like an eventuality with what seems to be a point of culmination, but lacking both preparatory phase and consequent state (could be marked as $\langle \emptyset, b, \emptyset \rangle$) cannot be interpreted as a telos. In other words, a triad having a consequent state implies that the triad also has a culmination point. However, the three respective components seemingly appearing on their own may easily be interpreted as corresponding to the notion *process* and *state* as used by Vendler and to the Bachian *point* expression.

Although the three non-complex eventualities (process, point, state) are not discussed further by Moens&Steedman, we deal with them in HuWN, and follow the above convention of showing the aspectual information in an ordered triple. Accordingly, the above listed possible combinations of the nucleus-components, each standing for one possible aspectual verb-subtype, are illustrated with examples, as follows:

¹⁴The sign \emptyset refers to non-conceptualised components of the triad.

$\langle \emptyset, \emptyset, \emptyset \rangle$	no example
$\langle a, b, c \rangle$	<i>befelhősödik</i>
$\langle a, \emptyset, c \rangle$	no example
$\langle \emptyset, b, c \rangle$	<i>eltörök</i>
$\langle a, b, \emptyset \rangle$	no example
$\langle a, \emptyset, \emptyset \rangle$	<i>fut</i>
$\langle \emptyset, b, \emptyset \rangle$	<i>kattan</i>
$\langle \emptyset, \emptyset, c \rangle$	<i>szeret</i>

Three of the possible combinations are excluded based on epistemologic grounds: (i) A nucleus having no components at all cannot be discussed neither conceptually nor linguistically. An eventuality (ii) having a preparatory phase and a culmination point, as well as one (iii) having a preparatory phase and a consequent state cannot be lexicalised due to the coherence of the telos and the consequent state.

Besides the remaining five lexicalised possibilities of nucleus-component combinations we have, however, seen the need for marking a sixth possible aspectual type in HuWN. As mentioned above, in many cases linguistic tests in Hungarian are unreliable in the sense that they provide ambiguous results even for native speakers. For the sake of usability in Hungarian language technology applications we considered it necessary to explicitly mark those cases in HuWN where the Hungarian test for the progressive did not result in a clearly grammatical sentence, but the English equivalent did. One such example can be seen in (5):

5. *János éppen gyógyult meg, amikor huzatot kapott a füle és újra belázasodott.*
John was getting better when his ear caught cold and he got fever again.

In cases like the above mentioned we decided to mark the first component of the nucleus "unmarked", designating this with an x: $\langle x, b, c \rangle$

3 The notion of the nucleus in HuWN

As we have seen, the conceptual presence or absence of meta-language elements beyond the lexicalized expressions can be tested with the help of Moens & Steedman's nucleus structure. The number of components a verb conceptualizes compared to an idealized complex event unit provides information on the telicity or atelicity of a given eventuality. If the third component of a nucleus denoted by a given verb is expressed,¹⁵ the eventuality is telic, if this component is not present, the eventuality is atelic.

3.1 Representing telicity in HuWN

From the six mentioned possible patterns whose lexicalisation the presence of the respective nucleus-components enables it is only complex eventualities that can be

¹⁵As mentioned in the previous section, the presence of the third component entails the presence of the second component.

telic. If we would like to get an overview of these complex eventualities from an aspectual point of view, the representation in ordered triples as introduced in 2.3, seems appropriate, as it can be seen in Table 2:

Table 2: Telicity of complex eventualities illustrated by the tripartite event structure of Moens&Steedman

Components of the triad	The metalinguistic name for the conceptualised components of the phrase lexicalising the triad	Telicity of the VP
<a, b, c>	<i>kimegy (to exit)</i> : <TO GO TOWARDS THE DOOR, TO STEP THROUGH THE DOOR, BEING OUTSIDE>	+consequent state→telic
<∅, b, c>	<i>felrobban (blow up)</i> : <∅, THE MOMENT OF BLOW-UP, BLOWN-UP STATE>	+consequent state→telic
<x, b, c>	<i>meggyógyul (get better)</i> : <x, CHANGING STATE FROM SICK TO HEALTHY, HEALTHY STATE>	+consequent state→telic

Of the simple eventualities, *processes* and *states* are usually considered atelic while *point* expressions (on their own, without context) are underspecified for this kind of information. When constructing a wordnet the question arises whether and how to represent meanings that should be synonyms according to the notion of synonymy in wordnet and yet differ aspectually. The notion of the nucleus helps us answer: aspectual differences can and should be represented in HuWN. If a meaning represented as a synset in the wordnet is transformed into a minimal proposition, one can determine whether the consequent state of the appropriate nucleus is present.¹⁶ By encoding whether a meaning has a consequent state (and hence a telos), through assigning to it one of the six conceptualization patterns of the triad components, the telicity of the eventuality expressed by the verb will be made explicit. This information is stored in HuWN in a similar way as in the case of the information on verb frames: we indicate which of the three triad components is conceptualised in Hungarian on the level of the literals.

As already introduced, for the sake of uniformity and transparency we follow the convention of showing the aspectual information in an ordered triple even in the case of simple eventualities mentioned in 2.2 and 2.3. Accordingly, the ordered triple of the verb *fut* 'run' is (< a, ∅, ∅ >). This triple shows on the one hand that the eventuality expressed by the verb *fut* is atelic, and on the other that it is

¹⁶Transforming verbal meanings into minimal propositions is ensured in the WordNet by mapping all the possible verbal subcategorisation frames of a given literal onto its synset. Sometimes several verb frames are merged into one verb frame record with optional arguments. In this case verbs should be considered with the minimal number of obligatory arguments. E.g. the verb frame *eszik*, 'eat' contains an optional direct object, so the minimal predicate should be formed without an object, and that predicate is atelic.

a Vendlerian process, indicated by the preparatory phase being solely present.

3.2 Complex eventualities in HuWN

Besides the possibility of storing a minimal amount of aspectual information concerning the given literal in a verb synset, the relational structure of the wordnet and the nucleus taken as a single unit allow us to propose another extension to the verb synset structure. In the case of complex eventualities whose certain triad components are not only conceptually present, but are *lexicalised*, as well, the unity of these components can be represented. Although the structure of PWN is based on a hierarchical system, an alternative structure has already been accepted for adjectives in PWN. By analogy it should be possible to organise the verb synsets in a slightly modified way than nouns, as well. The tripartite structure described above may be mapped onto the system of wordnet in the form of relations. The meta-language level described by Moens&Steedman's nucleus structure can be mapped onto the level of lexicalised elements, represented by wordnet synsets. The connection of the two levels is shown in Figure 4.

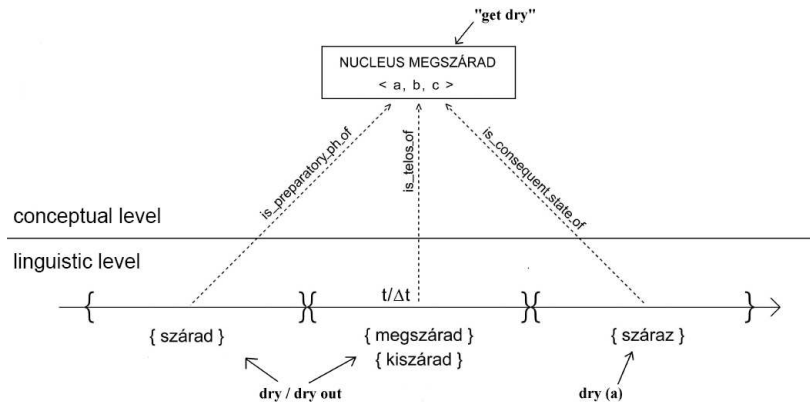


Figure 4: Applying the event-nucleus of Moens&Steedman to the synsets of WordNet

Artificial nodes introduced in HuWN (see [6]) are suitable for naming meta-language nuclei, e.g. the complex eventuality denoting the change of state from wet to dry, in the above example.¹⁷

The relational structure of the wordnet allows introducing three new relations according to the respective triad-components being related to the meta-language nucleus-unit, represented by an artificial node. These new relations point to the appropriate artificial node and they are called *is_preparatory_phase_of*, *is_telos_of* and *is_consequent_state_of*, respectively, based on the names of the different nucleus components.

¹⁷Artificial nodes are written with capital letters to distinguish them from natural language synsets.

Meanings that are lexicalized by a single verb in English but not in Hungarian can thus be distinguished: the same meaning might be present in Hungarian often as a verb with a preverb providing more aspectual information and as a verb without a preverb, more underspecified for aspectual information. In the above example, the Hungarian *szárad* and *megszárad* synsets are both equivalent to the English {dry:2}¹⁸. Without integrating the nucleus system into the wordnet the synset *megszárad* could be placed into HuWN only as a hyponym of *szárad*, considering all the originally available relations. However, this kind of storage would not distinguish the different implicational relation between the above mentioned two meanings, but would merge them into a hyponym-hypernym relation.¹⁹ After having integrated the nucleus system into the wordnet, there is no need for an additional explicit relation between the components of a nucleus: they are already connected through the artificial node. Following the path of the relations *is_preparatory_phase_of* and *is_telos_of*, it is easy to determine that the synset *szárad* represents the preparatory phase of the nucleus whose another lexicalized component is *megszárad*, hence *megszárad* implies *szárad*, while the implication does not hold in the other direction²⁰.

As we have seen, verbs belonging to the same triad (often with and without a preverb respectively) can be placed more accurately in HuWN with the help of the new relations. Furthermore, the relation *is_consequent_state_of* is not restricted to verbs, the third component of the triad mentioned above is the adjective synset *száraz* ({dry:1}). This psycholinguistically relevant piece of information is present in HuWN but would be lost if we had strictly held onto the structure of PWN without the tools for representing triads.

3.2.1 Triads in HuWN

Given that the presence of all the three triad-components presupposes an eventuality with a preparatory phase, a telos and a consequent state, we set out from the assumption that a possible domain where the adaptation of the nucleus-structure to the wordnet would be likely to prove useful would be that of verbs denoting some kind of change (e.g. change of state). Accordingly, we have chosen the unique beginner synsets²¹ {változik:1} ({change:1}) and {változtat:1} ({change:2}) — whose hyponyms alone make up at about the fourth of the verb synsets in the WordNet — to test the adaptation of the nucleus-structure to HuWN. When encoding the presence of the nucleus-components, the assumption that the hyponym trees of the above two unique beginner nodes would bear several synsets that lend themselves to being represented in a nucleus-structure proved to be right. Our results are

¹⁸*szárad* 'is drying' (v), *megszárad* 'get dry' (v), *száraz* 'dry' (a)

¹⁹By analogy to the nominal hypernymy relation, one way of conceiving of this relation between verbs would be basing it on selectional restrictions. E.g. the synsets {hervad, fonnyad} (fade, wither) and {rohad} (rot) would have such an ideal hypernymy relation, since the former selects plants as subject, while there is no such restriction on the subject of the latter one.

²⁰See Section 2.3 for a discussion on the connection (called contingency by Moens&Steedman) between the components of a triad.

²¹Synsets with no hyponyms are called *unique beginners* in WordNet terminology.

shown in Table 3.

Table 3: Proportion of nuclei under the synsets {változik:1} and {változtat:1}

nucleus type	{változik:1}	{változtat:1}
< \emptyset , b, c > or < x, b, c >	24 %	14 %
< a, b, c >	30 %	42 %
Altogether	54 %	56 %

After examining around 150 direct hyponyms of the mentioned two unique beginner synsets we have found that in more than half of the cases the application of the nucleus-structure facilitated the positioning and representation of a meaning lexicalised in Hungarian in the network.

The conceptual similarity between the two unique beginner nodes could lead to the assumption that the triads associated with them have a parallel internal structure, too. However, at a closer examination one sees that the relations between the respective components of the nucleus differ as a result of the causative-inchoative alternation. The difference may be explained through the following example: The second and third component of the nucleus <SZÁRAD, MEGSZÁRAD, SZÁRAZ> are connected by a consequent state relation, since the consequent state of the event expressed by the verb *megszárad* is the dry state of its subject. The seemingly similar structure of the nucleus <SZÁRÍT, MEGSZÁRÍT, TULVAN A MEGSZÁRÍTÁS FOLYAMATÁN> hides a different relation between the second and the third component: the consequent state of the eventuality refers to the object of the verb *megszárít*, meaning that the synset {megszárít} has a *causes* relation to the adjective synset {száraz} (dry). Although the *causes* relation appears in the English WordNet between the nodes {dry:1} (causative) and {dry:2} (inchoative)²², there is no relation that would connect the verb synset {dry:2} with the adjective {dry:1}. If the Hungarian WordNet was to follow the relation patterns of PWN, a *causes* relation would have to be encoded between both causative-inchoative verb-pairs (*szárít-szárad* and *megszárít-megszárad*). However, this would unnecessarily duplicate the number of relations encoded, while the relation between the causative verbs and the cause itself, expressed by the adjective synset, would still only be indirectly visible. This is why we propose to connect the metalinguistic nodes denoting the triads with a *causes* relation (i.e. *megszárít-causes→megszárad*), which allows for interpreting the *causes*-relation between the verbs of the causative and the inchoative triad as well as between the verbs of the causative triad and the adjective synset expressing the consequent state of the inchoative triad.

²²Unfortunately the encoding of these relations is not systematic in the PWN.

4 Possible applications

Besides the fact that one of the main tasks of a wordnet is to provide a uniform representation for the idiosyncratic properties of the lexical items, the extension of HuWN in the proposed way brings practical benefits, as well. As we have seen, it can be easily deduced from the triad whether a given verb is telic or atelic, perfect or progressive, respectively. A Hungarian-English MT system can be improved by using this information provided in the HuWN, e.g. in the area of matching the verb tenses in the source and the target language more appropriately. Since there are only two morphologically marked tenses in Hungarian (present and past), a rule-based MT system would select the same two tenses in the target language, simple present and simple past, respectively. Inaccurate translations would emerge inevitably. However, the above outlined information integrated into HuWN would improve the system. In Hungarian, for example, morphologically present tense forms of a telic verb have a future reference. The English equivalent of the Hungarian sentence *Felhívom Pétert* is not *I call Peter*, but *I will call Peter*. Similarly, progressive past tense verb forms should be matched with the past continuous form of the appropriate verb, instead of selecting the simple past form: the Hungarian *Péter az udvaron játszott* should be matched to the English *Peter was playing in the yard*, instead of the now expected *Peter played in the yard*. Aspectual information may be used in generating sentences, as well, whether it be translation from English to Hungarian, or some other tasks requiring generation.

These properties of verbs may be helpful in machine comprehension, as well. The knowledge of such idiosyncratic properties of verbs is an important component of the inner representation of a computer. Without this information, just by considering the temporal adverbials (possibly) present in the sentence, it is not possible to represent or reconstruct the temporal structure of a narrative accurately.

5 Conclusions

In the present paper we have tried to show on the example of the Hungarian WordNet in what ways the wordnet-structure as conceived of in PWN may be exploited and extended in order to represent some language-specific and part of speech specific phenomena of typologically different languages than English, as well. Although specifically implemented for solving a linguistic situation in the Hungarian language, the implementation of the nucleus-structure in the WordNet in the form of relations might prove to be useful for other languages with a rich morphology showing aspectual distinctions, as well. Later applications of the extended HuWN will hopefully prove the direction in which we tried to point with the above detailed representation of verbs to be useful.

References

- [1] Bach, E. The Algebra of Events. *Linguistics and Philosophy*, 9: 5-16, 1986.
- [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] Kiefer, F. *Aspektus és akcióminőség. Különös tekintettel a magyar nyelvre*. Akadémiai Kiadó, Budapest, 2006.
- [6] Kuti, J., Vajda, P. and Varasdi K. Javaslat a magyar igei WordNet kialakítására. In Alexin, Z. and Csendes, D., editors, *III. Magyar Számítógépes Nyelvészeti Konferencia*, 79–88, Szeged, 2005. Szegedi Tudományegyetem.
- [7] Moens, M., Steedman, M. Temporal ontology and temporal reference. *Computational Linguistics*, 14(2):15–28, 1998.
- [8] Tufis, D. et al. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *Romanian Journal of Information Science&Technology*, 7(1-2):1–35, 2004.
- [9] Vendler, Z. Verbs and Times. *Philosophical Review*, 66:143–160, 1957.
- [10] Verkuyl, H. J. *On the compositional nature of the aspects* Foundations of Language, Supplementary Series, 15, Reidel, Dordrecht, 1972.
- [11] Vossen, P. EuroWordNet General Document. Technical Report EuroWordNet (LE2-4003, LE4-8328), 2005.