Definitions of Lexical Meanings: Some Reflections on Purpose and Structure

Leo Wanner
Intelligent Systems Institute
University of Stuttgart
Universitätsstr. 38, D-70569 Stuttgart
wanner@informatik.uni-stuttgart.de

Résumé - Abstract

Dans le cadre de la Théorie Sens-Texte, la décomposition du sens est un des principes fondamentaux du modèle linguistique sous-jacent. Chaque entrée du dictionnaire explicatif et combinatoire d'une langue $\mathcal{L}$ contient la spécification du sens du vocable en question en termes d'une définition écrite dans le langage restreint de $\mathcal{L}$. À cause de leur lisibilité pour les lecteurs humains, les définitions en langage naturel sont difficiles à traiter par des ordinateurs – que ce soit dans le but de maintenir la cohérence du dictionnaire ou pour leur emploi dans des applications de traitement automatique des langues naturelles. Nous défendons l'idée que (i) les définitions lexicographiques doivent comporter une structure détaillée afin d'accepter les opérations sur le sens lexical par les diverses applications de TAL et (ii) les définitions doivent être rigoureusement formalisées. Il n'est pas très difficile de calculer automatiquement la forme textuelle d'une définition formelle structurée, mais la réciproque est extrêmement difficile.

In the Meaning-Text Theory, meaning decomposition is one of the fundamental principles of the underlying linguistic model. Each entry in the *Explanatory Combinatorial Dictionary* of the language $\mathcal{L}$ contains the specification of the meaning of the lemma in question in terms of a definition written in restricted $\mathcal{L}$. Being suitable for human readers, natural language definitions are hard to process by computers—be it for purposes of dictionary consistency maintenance or for their use in Natural Language Processing applications. We argue that (i) lexicographic definitions must receive a detailed structure to support the operations on lexical meaning by different NLP-applications, and (ii) definitions must be rigorously formalized. It is not so hard to automatically compute the textual form of a formal and structured definition, but it is very hard to compute a formal and structured definition from a textual form.

Mots-clefs – Keywords

dictionnaire, définition, décomposition du sens, structure de la définition, TAL
lexicon, definition, decomposition of lexical meaning, structure of the definition, NLP

1 Introduction

Most of the modern semantic theories used in *Natural Language Processing* (NLP) presuppose a non-decompositional description of the meaning of lexical units (LUs). As a result, they deal with "shallow" semantic representations which feature an obligatory one-to-one relation between a *signified* and its *signifier*. In contrast, in the *Meaning-Text Theory* (MTT) meaning decomposition is one of the fundamental principles of the underlying linguistic model. The meanings of lexemes are given in a decomposed form in the lexical component of the theory, the *Explanatory Combinatorial Dictionary* (ECD), in a textual format—the claim being that it is only when the meaning of a lexeme is decomposed that its syntactic and semantic valency as well as its restricted lexical co-occurrence can be made explicit. However, decompositions as usually given so far are tailored towards human users, not towards the machine. Consider, for illustration, the definition of Fr. BERCER '[to] cradle' from (Mel'čuk, 1999, 116):

> *X BERCE Y*: Personne X cause que la personne Y se balance doucement et régulièrement, dans le but d'apaiser ou d'endormir Y.

The part of the definition given in italics (called *propositional form*) specifies the semantic argument structure of the lexeme, the rest of the definition (we call it the *textual form*) specifies its decomposed meaning.

In papers on *Explanatory Combinatorial Lexicology* (ECL), one often finds two statements: **1.** The textual form is the linearized counterpart of the formal semantic network that encodes the meaning of the lexeme in question, and can (easily) be transformed into it.[1] **2.** The definition of a lexeme can replace its signified in any well-formed semantic structure, the replacement resulting again in a well-formed structure.

However, to our knowledge no formal outline, let alone an implementation, of how both operations are to be realized exists. We are convinced that this is not by chance: (i) implies all aspects of "deep parsing"—including all problematic ones; (ii) implies a subtly differentiated handling of the various parts of the definition. Therefore, we think that in order to be suitable for NLP, the ECD-like definitions must be given a well-defined structure and the statements in each substructure must be formalized. While the need for formalization can be assumed to be undisputable, the need for an additional structuring of the definition is often received with certain scepticism—although some proposals for different structures of ECD-definitions have already been discussed; cf. (Iordanskaja & Mel'čuk, 1990; Mel'čuk & Wanner, 2001), and also (Polguère, this volume). In what follows, we attempt to provide more evidence for such a need. Due to the lack of space, we cannot go into details. Nonetheless, we hope that our argumentation will ignite a discussion on this important topic.

In Section 2, we argue that meaning decomposition is an absolute must for advanced NLP, and the granularity of meaning decomposition as proposed in MTT is necessary. Then, in Section 3, we review the proposals for structuring definitions known from the MTT-literature. In Section 4, we analyse some of the requirements with respect to the structure of definitions as put forward

[1] The preference of the textual notation over the network notation is justified by the argument that it is more convenient for presentation in a dictionary and that humans can easier encode and decode textual forms; cf., e.g., (Mel'čuk *et al.*, 1995, 73*f.*).

by NLP-applications and suggest a structure for them that is more detailed and more formalized, and thus more suitable for NLP than present proposals. In Section 5, finally, we summarize our own proposal.

2 On the Relevance of the Decomposition of Meaning in NLP

Before embarking upon the difficult and controversal issue of how to structure and formalize lexical definitions, we must be sure that meaning decomposition as such is really needed. A look at any advanced NLP-application shows that meaning decomposition is indispensable. In Text Generation we need meaning decomposition, for instance, for lexicalization and paraphrasing. Thus, we need to know that the predicate structure 'go(John,Boston,plane)' can be realized as *John flies to Boston*, i.e., that the content 'use plane as means of transportation' matches the meaning of the word FLY, while the content 'use train as means of transportation' does not match the meaning of any single word (and we must thus realize 'go(John,Boston,train)' as *John goes to Boston by train* or *John takes the train to Boston*). When we report on the difference of the temperature t at time T_1 and time T_2, and we know that $t(T_1) < t(T_2)$, it is only when we have the following information in the dictionary that we are in a position to choose the verb [*to*] *rise* to name the situation:

> *X rises*: X has the value v at time t and the value v' at time $t + \Delta$, such that $v < v'$.

In Machine Translation, the need for meaning decomposition is most clearly seen in the case of lexical mismatches (including collocations). Lexical mismatches and meaning decomposition in connection with lexical mismatches are discussed at length in (Mel'čuk & Wanner, 2001). To see the relevance of meaning decomposition for the translation of collocations, consider some collocates of the German noun KAMPF and some collocates of its English equivalent FIGHT.[2]

KAMPF:

> *aussichtslos* lit. 'with no prospect' ('hopeless', 'desperate', 'pointless'), *erbarmungslos* 'pitiless'/'merciless', *erbittert* 'bitter', *fair* 'fair', *heftig* lit. 'severe' ('vehement', 'intense', 'fierce', 'hard', 'violent'), *schwer* lit. 'heavy' ('desperate', 'fierce', 'hard'), *tödlich* lit. 'mortal', *unbarmherzig* 'pitiless'/'merciless', *verbissen* lit. 'bitten-into' ('grim', 'dogged', 'determined'), *verzweifelt* 'desperate'

FIGHT:

> *bitter, desperate, fierce, hard, stubborn, clean, dirty, last-ditch*

With a sketchy definition of KAMPF 'fight' being:

Der Kampf von X gegen Y für/gegen Z 'The fight of X against Y because of Z':

> Handlungen von X, die er gegen den Widerstand von Y unter Aufwendung großer Anstrengungen und mit gewisser Beharrlichkeit unternimmt, um Z, das für Y oder irgend ein α negative Auswirkungen hat, zu erreichen bzw. zu verhindern.
>
> lit. 'Actions of X which he carries out against the resistance of Y, making great efforts and with a certain persistence in order to achieve respectively avoid Z which has negative consequenzen for Y or for any other α.'

[2] The collocates of FIGHT are from (Benson *et al.*, 1986).

When considered in isolation, *verzweifelt* and *desperate* are translation equivalents. Therefore, we might conclude that this will be still the case when they appear in collocations—e.g., in *verzweifelter Kampf* and *desperate fight*, respectively. The same applies, e.g., to *erbittert* and *bitter*.[3] However, for *verbissen*, *unbarmherzig*, and *heftig*, no literal translation equivalents are available in the collocate set of FIGHT. Therefore, only if we know that, e.g., *stubborn* correlates with the same meaning component in the definition of FIGHT as *verbissen* does with respect to the definition of KAMPF (namely 'persistence', respectively 'Beharrlichkeit'), while, e.g., *bitter* correlates with the component 'resistance', and *last-ditch* correlates with the component 'negative consequences', we are able to correctly translate *verbissener Kampf* as *stubborn fight*. Obviously, to be able to identify the proper meaning components, we must decompose the meanings of KAMPF and FIGHT.

Other NLP-applications, such as automatic acquisition of collocations, text summarization, analysis, and so on equally require meaning decomposition.

3 Structure of Lexical Definitions in ECDs: A Review

When we talk about the structure of lexical definitions, we mean a *block* (or *standardized schema*) structure where each block contains a part of the definition that is distinct in its nature from the other parts. The idea of a block structure for definitions is well-known in lexicology; see, among others, Apresjan (1974), Wierzbicka (1985), Glovinskaja (1993), Iordanskaja and Mel'čuk (1990), Pustejovsky (1995), and Paducheva (1997). The most popular definition block structure in Computational Lexicology (CL) is the one introduced by J. Pustejovsky in the *Generative Lexicon Theory*. It would be worthwhile to analyze Pustejovsky's structure from the perspective of MTT. Unfortunately, due to the lack of space, we must refrain from such an analysis as well as from an adequate review of the state of the art in the area of definition structuring. In what follows, we focus on the format of lexical definitions in ECL only.

According to Mel'čuk *et al.* (1995), a definition is composed of the following blocks:

1. A block that contains the *standard components* of the definition.
2. A block that contains the *generic component* of the definition, which specifies the class to which the lexical item belongs.
3. A block with *weak components* of the definition, which are eliminated in certain semantic contexts. For instance,[4] in the definition of Germ. STUDENT, the component 'männlich' (= 'male') is weak: it must be eliminated, when STUDENT is used in plural or in the generic sense:[5]

   (1) a. *Die Studenten protestieren gegen die Studiengebühren*
   lit. 'The students protest against the tuition fees'.
   b. *Der Student von heute steht nicht gerne früh auf*
   lit. 'Today's student does not like to get up early'.

[3] Note, however, that this type of inference does not always hold.

[4] The example is the German translation of the example from (Mel'čuk *et al.*, 1995, 95).

[5] For the argumentation, why 'male' must form a component in the definition of STUDENT, see (Mel'čuk *et al.*, 1995, 93*f.*).

4. A block with optional components of the definition, which can be neutralized by contradicting components in specific contexts. For instance, the component ‘in a public institution or enterprise’ in the definition of SEKRETÄRIN is neutralized by the context ‘private’ in an utterance like

   (2) *Frau Müller ist die Privatsekretärin des alten Fürsten*
   lit. ‘Frau Müller is the private secretary of the old duke’.

5. A block with *restrictions* on the internal variables. These can be both type and attributive restrictions (as, e.g., ‘type:person’ for the first two actants in the definition of LEHRER ‘teacher’).

6. A block with *presupposed* components of the definition, which remain asserted when the whole meaning is negated. Thus, consider the component ‘female’ in the definition of the lexeme SCHWANGER ‘pregnant’: if such a statement as *Sie ist schwanger* ‘She is pregnant’ is negated, it remains stated that the person in question is female.

To indicate the structure of the definition, Mel’čuk *et al.* propose to emphasize certain blocks by a different font, to use specific syntactic constructions for certain blocks, and to put the blocks in a certain order. Obviously, none of these means is suitable for making the structure more accessible in NLP-applications.

So far, a few case studies have been carried out in ECL on explicit definitorial block structures—among them (Iordanskaja & Mel’čuk, 1990) and (Mel’čuk & Wanner, 2001). Iordanskaja and Mel’čuk consider emotion lexemes in Russian, Mel’čuk and Wanner speech act nouns in German. The two studies show that the structure depends on the type of the lexeme concerned. This is in contrast to, e.g., the view adopted by Pustejovsky and his followers, who propose to use the same structure for all types of lexemes. The structure proposed by Iordanskaja and Mel’čuk for emotion lexemes is cited in Figure 1.

X boitsja Y-a ‘X is afraid of Y’

Presupposition: —
Assertion:
Evaluation: ‘X expects that Y will cause something undesirable for X
or
Emotion (of X): X is in or has the property of being in
Characterization: an unpleasant emotional state with respect to Y,
Cause: which is caused by said X’s expectation,
Prototyp.Conditions: this state being such as is usually caused by the expectation of something dangereous,
Effect: this expectation and/or state causing X to tend to avoid Y’

Figure 1: Structured definition of BOJAT′SJA ‘to be afraid’ as proposed by Iordanskaja and Mel’čuk

The two main blocks of their structure are Presupposition and Assertion. The Assertion-Block is further subdidvided into an *Evaluation*-, an *Emotion*- and an *Effect*-subblock. The *Emotion*-subblock contains the main characteristics of the emotion lexeme described and the actual and prototypical conditions.

AUFRUF(X,Y,Z) 'appeal'

Generic component:
text, meant to be publicly accessible, by means of which

Assertion:
Core Action: X tries to cause Y to do Z
Speech Act: by communicating to Y that X wants Y to do Z
Purpose (of speech act): in order to transmit X's desire
to Y and cause Y to want do Z
Characteristics:
1. Z is specified sufficiently for Y to understand it
2. (communicating is emotional)

Presuppositions:
1. X believes that γ, where
γ := {1.1 there is a state of affairs α that concerns X,
Y or someone else
1.2 α requires Z
1.3 Y can and should do Z}

Actants:
X = Communicator: official OR official body
Y = Addressee: collectivity
Z = Goal: action OR activity

Figure 2: Structured definition of AUFRUF 'appeal' as suggested by Mel'čuk and Wanner

Mel'čuk & Wanner (2001) suggest a structure for speech act nouns cited in Figure 2.

The structure in Figure 2 follows the outline in Mel'čuk *et al.* more closely than the one in Figure 1 does. Four main blocks are distinguished: in addition to the Presupposition and Assertion blocks already used by Iordanskaja and Mel'čuk, a Generic block and an Actants block are introduced.

The main purpose of block structures shown in Figures 1 and 2 has been to facilitate the comparison of definitions first of all by a human reader. Therefore, these block structures keep, at least partially, their sentential form. However, to facilitate NLP-tasks, a definitorial block structure must be even more detailed. Furthermore, it must avoid the necessity of parsing a definition in order to filter out information needed for a specific task, i.e., it must be more formal and easier to access.

4 Towards an NLP-Suitable Block Structure of Definitions

In connection with NLP, the two important questions to be asked with respect to meaning decomposition are: (i) which tasks must the decomposition of the meaning of a lexeme facilitate?, (ii) how must the structure of the decomposed meaning look like in order to support these tasks? Let us discuss these two questions in turn.

4.1 How Can Definition Structure Facilitate NLP-Tasks?

Definitions are complex elements of the formal language used for the specification of linguistic representations at the semantic level. Therefore, they must facilitate: (i) the well-formedness check of the semantic structure in which the signified of the lexeme in question occurs, (ii) the substitution of the signified of the lexeme in question in a semantic structure by its definition, (iii) the substitution of the signified of the lexeme in question in a semantic structure by a suitable fragment of its definition, (iv) the generalization of components across definitions of different lexemes, (v) the relation of collocates of the lexeme in question to the meaning components they correlate with, (vi) the comparison of the definitions of several lexemes, (vii) the well-formedness check of the definition itself.

Due to the lack of space, we discuss, in what follows, only the first three of these tasks.

4.1.1 Facilitating the Well-Formedness Check of the Semantic Structure

The semantic structure S_{sem} in which a semanteme σ_i occurs is not well-formed if it contradicts to one or several components in the definition of the meaning of σ_i. Thus, we immediately see that the semantic structure of *He leant on the back of the stool* is not well-formed because the definition of STOOL contains the component 'with no back'. However, due to the possibility of weak and optional meaning components in a definition, a more differentiated formulation of the well-formedness criteria is necessary: In order to be well-formed, a semantic structure S_{sem} must not contradict: (i) the type of the actants of any σ_i that occurs in S_{sem}, the obligatory non-weak (i.e. the *strong*) components in the definition of the signifier of any $\sigma_i \in S_{sem}$.

That is, S_{sem} may well contain a contradiction with a weak or an optional component of the meaning of $\sigma_i \in S_{sem}$. Consider two examples:

(3) the weak component 'male' in the definition of STUDENT 'student' is removed in the present context:

Maria und Hans sind gute Studenten 'Maria and Hans are good students'.

(4) the optional component 'of red, yellow or green color' in the definition of APFEL 'apple' is substituted in the present context by 'purple':

Maria schenkte mir einen lila Apfel 'Maria gave me as present a purple apple'.

This means (in accordance with Mel'čuk *et al.*) that the following components must be singled out in the structure of a definition: (i) the type specifications, (ii) the weak and the strong components of the definition, (iii) the obligatory and the optional components of the definition.

Some of the components that fall into one of these categories are very prominent for well-defined classes of lexemes, and should thus be specified in separate subblocks. An example of such a component is the specification of the temporal structure of a functor that falls into the category of strong (or "standard" in I. Mel'čuk's terms) components. The contradiction of the context of a functor within a S_{sem} with the temporal structure of this functor leads to a not well-formed structure; cf.:

(5) a. **Hans begann den Stein zu <u>werfen</u>* 'John began to throw the rock'.

b. **Hans fing an, <u>krank</u> zu sein* 'John started to be ill'.

c. **Hans hörte auf zu sterben* 'John stopped to die'.

The introduction of a temporal structure subblock (*event structure* in Pustejovsky's *Generative Lexicon Theory*) is thus well-justified.

4.1.2 The Substitution of the Signified by Definition

The substitution of the signified of a lexeme σ_L by its definition $\mathcal{D}_L$ in a semantic structure S_{sem} consists, roughly speaking, of two steps: (i) the replacement of σ_L by $\mathcal{D}_L$, and (ii) embedding of $\mathcal{D}_L$ in $S_{sem}\backslash\sigma_L$, i.e., the connection of selected nodes in $\mathcal{D}_L$ with the nodes with which σ_L was connected in S_{sem}.

In order to facilitate the replacement, the head of $\mathcal{D}_L$ must be singled out. As far as the embedding is concerned, two main cases need to be distinguished: **(1)** σ_L is a functor with one or several instantiated actants in S_{sem}. In the case of substitution, these actants become the actants of the head of $\mathcal{D}_L$. The unification of the actant variables in $\mathcal{D}_L$ with the actants of σ_L requires the explicit specification of the argument structure of the head of $\mathcal{D}_L$. **(2)** σ_L is itself an actant of a functor in S_{sem}. Case (2) subdivides into three subcases: (2.1) σ_L is an actant of an action, state, event, or connector, (2.2) σ_L is an actant of a property or circumstantial, (2.3) σ_L is an actant of a negation. No further structuring of the definition needs to be done to account for **(2.1)**. If σ_L is an actant of a property σ_P (subcase **2.2**), σ_P may be related either to the head of $\mathcal{D}_L$ or to a dependent of the head of $\mathcal{D}_L$. Thus, consider the definition of Fr. BERCER in Section 1 and sentence (6):

(6) *La grand-mère berce le petit Jean dans ses bras* 'The grandmother cradles little Jean in her arms'.

If we replace in the semantic structure of (6) 'bercer' by the definition of BERCER attaching all relations of 'bercer' to 'causer' (the head of the definition), we get something like (7a) (instead of the correct 7b):

(7) a. **La grand-mère cause dans ses bras que le petit Jean se balance doucement et régulièrement, dans le but de l'apaiser ou de l'endormir.*

b. *La grand-mère cause que le petit Jean se balance dans ses bras doucement et régulièrement, dans le but de l'apaiser ou de l'endormir.*

That is, *dans ses bras* must be related to *se balancer*, which is a dependent of *causer*, not to *causer* itself. Therefore, in the definition for BERCER, 'se balancer' must be singled out as the "secondary head" which serves as the target node of certain relations.

In the subcase **(2.3)**, σ_L is negated in S_{sem}—as, e.g., 'pregnant' is in the semantic structure of

(8) *Mary is not pregnant.*

(We assume that the semantics of this proposition is 'Mary is in the state of non-pregnancy' rather than 'Mary is not in the state of pregnancy'.)

A sketchy definition of PREGNANT would be

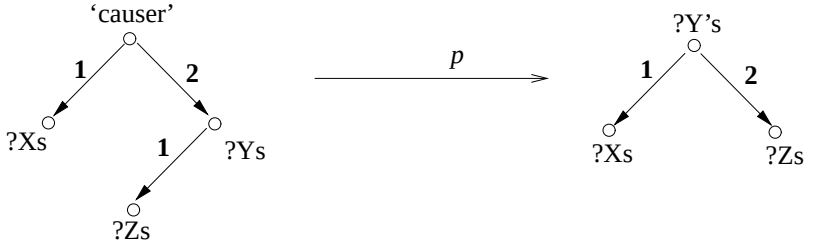
X is pregnant: A female mammal X carries a fetus in her belly.

The negation of the definition of PREGNANT in the S_{sem} does not affect the component 'female mammal': it must remain affirmed. But when we replace *woman* in the structure of *This individual is not a woman* by the definition of *woman*, 'female' must be negated (while 'person' must remain affirmed). In other words, we need an explicit distinction between 'presupposed' and 'asserted' parts of a definition. However, this distinction should not form the upper level of a bipartite structure as suggested by Iordanskaja and Mel'čuk above: 'presupposed' vs. 'asserted' is only one among several block distinctions, and the question on whether a certain block is asserted or presupposed depends on the type of the lexeme in question.

4.1.3 The Substitution of the Signified by a Part of the Definition

A definition of a lexeme is not thought to be used as a valid paraphrase of this lexeme in the lexical choice process. But, as has often been argued, the generic component of the definition does provide such a paraphrase. This is another reason to introduce a separate block for the generic component. However, it is not only the generic component that can be used in paraphrasing. Thus, in the definition of BERCER, the fragment 'causer que la personne Y se balance doucement et régulièrement' is a valid semantic paraphrase of BERCER after a semantic rule as sketched in (9) has been applied to it.[6]

(9) *La grand-mère balance doucement et régulièrement le petit Jean dans ses bras.*



That is, in order to be able to use the above fragment as a paraphrase of BERCER, we have to single it out in the definition, and we have to indicate that the rule in (9) must be applied to it.

As shown by J. Milićević, periphrastic substitutions of lexemes by fragments of their definitions are very common. Consider, for illustration:

'X flies'	$\sim$	'X travels by plane'
'X expects Y'	$\sim$	X believes that probably Y'
'X buys Y'	$\sim$	'X becomes owner of Y'

Most often, a semantic paraphrase requires the inclusion of the generic component (see, however, the last of the above examples). But, as the examples show, in addition to the generic component, further components make part of the paraphrase (up to the complete definition).

4.2 Towards a Prototypical Block Structure

Let us now put together a prototypical block structure that results from our discussion above. As already became clear, certain blocks that we consider as useful belong to the theoretical stock of

[6]For a detailed discussion of semantic paraphrasing rules, see (Milićević, in preparation) and (Wanner, in preparation). The rule cited is simplified and, strictly speaking, even incorrect. It does not distinguish between the context, the actual lefthand side of the rule that is mapped onto the righthand side, and the embedding statements. However, it suffices for purposes of illustration.

the blocks listed in (Mel'čuk *et al.*, 1995). Our argumentation thus provides further evidence for the necessity to identify these blocks when writing definitions. In addition to these "established" blocks, a number of further blocks that have not yet been discussed in the MTT-literature turn out to be useful.

Internal:

Head: $causer(X, balancer2(Y))$

Sec. Head: $\alpha : balancer2(Y)$

$\sigma_L \overset{1}{\leftarrow} < LIEU/TEMPS > \Rightarrow \alpha \overset{1}{\leftarrow} < LIEU/TEMPS >$

$\sigma_L \overset{1}{\leftarrow} < ATTRIBUTE > \Rightarrow \alpha \overset{1}{\leftarrow} < ATTRIBUTE >$

Temporal Structure: $\uparrow ACTIVITY \uparrow ATELIC$

Generic: $balancer1$

Actant Structure:

X: $personne$; Y: $personne$

Equivalence:

1: $causer(X, \alpha) \wedge doucement(\alpha) \wedge reguliement(\alpha) \parallel Prule_{ca}$

Logical:

Presupposition: –

Assertion:

Core: $causer(X, \alpha)$

Characteristics: $doucement(\alpha)$
$reguliement(\alpha)$

Goal (of X): $\beta : apaiser(Y), \gamma : endormir(Y)$

Secondary:

Weak: –; *Optional*: β, γ

Figure 3: A sketch of the structure of the definition of BERCER

Since, as is well known, the block structure of a definition depends on the type of the LU in question, it is not useful to attempt to provide a "universal" structure in which some parts are filled out and some remain empty—depending on the concrete case. That is, the block structure of concrete object names should differ from the block structure of functor names, and the block structure of predicates should differ from the block structure of quasi-functor names, i.e., object lexemes that possess an argument structure, as, e.g., Fr. PINCEAU, COUTEAU, CASSEROLE, LUNETTE, etc. Due to the lack of space, we cannot elaborate on the block structure of different types of lexemes. This must be left to a more detailed presentation of our work. In Figure 3, we simply summarize the blocks mentioned above and apply them to the definition of BERCER. For better readability, we refrain from a formal presentation and use a quasi-logical notation instead. 'σ_L' stands for "signified of the lexeme L in question", in this case BERCER. 'Balancer1' stands for the causative sense of BALANCER, while 'balancer2' for its non-causative sense. The 'Secondary Head' slot may also contain the rules that need to be applied in order to embed the definition into a semantic structure. Fragments of the definition that may serve as a paraphrase of the lexical item in question are summarized under the heading 'Equivalence'. If a rule is specified (as in this case), it must be applied to the fragment in question. We specify 'Presupposition' and 'Assertion' under the heading 'Logical' because it is applied here to negation only. 'Assertion' may receive a further hierarchical organization (as, e.g., in Figure 2). Recall that the structure in Figure 3 contains only blocks discussed in the course of the paper. The tasks mentioned in Section 2 but not discussed in the course of the paper require the introduction of additional blocks.

5 Conclusions and Future Work

Given that ECDs are designed for human use, it is well justified that traditional definitions in the sense of MTT possess a textual form with no further structure being superimposed on a definition. For NLP, the textual form is not suitable due to several reasons: (i) it requires the parsing of the definition each time an access to a specific aspect of the meaning of a lexeme is needed, (ii) it does not support certain operations such as paraphrasing, (iii) it always implies a certain amount of redundancy. Therefore, at least for NLP-applications, structured and formal descriptions of lexical meanings are needed.

In this paper, we have discussed some of the criteria that must be taken into account when designing the block structure of definitions. In order to provide comprehensive and well-motivated structures for different types of lexical items, further systematic studies must be carried out.

References

APRESJAN J. (1974). *Leksičeskaja semantika*. Moscow: Nauka. 2nd revised edition published 1995, Moscow: Škola "Jazyki russkoj kul′tury"; the English translation appeared as *Lexical Semantics. User's Guide to Contemporary Russian Vocabulary*. Ann Arbor: Karoma. 1992.

BENSON M., BENSON E. & ILSON R. (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins Publishing Company.

GLOVINSKAJA M. (1993). Semantics of Speech Act Verbs within the Framework of the Speech Act Theory. In E. ZEMSKAJA, Ed., *Russkij jazyk v ego funkcionirovanii. Kommunikativno-pragmatičeskij aspekt*, p. 158–218. Moscow: Nauka.

IORDANSKAJA L. & MEL'ČUK I. (1990). Semantics of Two Emotion Verbs in Russian: BOJAT′SJA 'to be afraid' & NADEJAT′SJA 'to hope'. *Australian Journal of Linguistics*, **10**(2), 307–357. Reprinted in Mel'čuk, I. 1995. *The Russian Language in the Meaning-Text Perspective*, Moscow/Vienna: Škola "Jazyki russkoj kul′tury"/Wiener Slawistischer Almanach, 81-124.

MEL'ČUK I. & WANNER L. (2001). Towards a lexicographic approach to lexical transfer in machine translation (illustrated by the German-Russian language pair). *Machine Translation*, **16**(1), 21–87.

MEL'ČUK I. A., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

MEL'ČUK, I. A. *et al.* (1999). *Dictionnaire explicatif et combinatoire du français contemporain*. Montréal, Canada: Presses de l'Université de Montréal. Volume 4.

MILIĆEVIĆ J. 2003. PhD thesis, Département de linguistique et de traduction, Université de Montréal.

PADUCHEVA E. (1997). Verb Categorization and the Format of a Lexicographic Definition (Semantic Types of Causative Relations). In L. WANNER, Ed., *Recent Trends in Meaning-Text Theory*. Amsterdam: Benjamins Academic Publishers.

PUSTEJOVSKY J. (1995). *The Generative Lexicon*. Cambridge: The MIT Press.

WANNER L. (in preparation). *Dekompositionelle Semantik in der Sprachverarbeitung*. Stuttgart: Universität Stuttgart.

WIERZBICKA A. (1985). *Lexicography and Conceptual Analysis*. Ann Arbor, MI: Karoma.