HOW MUCH GRAMMAR IS NEEDED IN LEXICON?

Anita Peti-Stantić
Filozofski fakultet, Zagreb

UDK 81'374:81'34'36'37–116.6

Čeprav lahko razumemo in sprejmemo razloge za osnovno razlikovanje med slovarjem in slovnico, razvoj stroke v zadnjih nekaj desetletjih 20. stoletja govori o področjih njunega prekrivanja in poenotenja. V tem prispevku nam začetno točko za analizo predstavlja teorija paralelne arhitekture. Gre za teorijo, ki določa jezikovno strukturo tako skozi tri neodvisne generativne sisteme – fonologijo, sintakso in semantiko – kot tudi skozi povezave med njimi. Za primere v tem prispevku sem izbrala glagole in samostalnike, saj sem hotela pokazati, da celo leksemi z bogato semantično strukturo vsebujejo precejšnje število slovničnih informacij že v svoji osnovni leksični obliki.

slovar, slovnica, paralelna arhitektura, generativni sistemi

Although one can understand and accept the reasons for the distinction between lexicon and grammar, the development of the field in the last quarter of the 20th century brought to our attention domains of their overlap and unification. The starting point for the present analysis is Parallel Architecture, a theoretical framework which takes linguistic structure to be determined by three independent generative systems – phonology, syntax, and semantics, as well as the links between them. Choosing lexemes from the class of nouns and verbs, I want to show that even lexemes rich in semantic structures carry a significant amount of grammatical information within their lexical form.

lexicon, grammar, Parallel Architecture, generative systems

1 Theoretical grounding

One can understand and accept the reasons for the principal distinction between lexicon and grammar. To some extent, this distinction is in line with the centuries-long practical activity of producing dictionaries, lexicons and grammar books, as well as with the quite recent distinction between levels of linguistic analysis. However, the development of the field in the last quarter of the 20th century brought to our attention domains of their overlap and unification.

Among the main lines of research that challenged this divide are contemporary psycholinguistic processing studies. They brought under scrutiny the assumption of a strict divide between the lexicon and the grammar, or the partition of knowledge between the lexical list (or, to express it in a more complex fashion, semantic net) and the rules of grammar that cannot be justified by analyzing linguistic phenomena in natural languages. Both production and comprehension studies proved to be a reliable testing ground for the relationship of the lexicon and the syntax, especially for their

mutual interdependence in a much more substantial way than the one anticipated by traditional generative grammar. Both syntactically and semantically driven models of language processing are in line with the psychological reality that every lexeme sets the frame for its place in the language system. The frame has to be defined in terms of the conceptual structure, as well as in terms of the inherent morphosyntactic features that place constraints on the licensing domain of every particular lexeme. After introducing a particular lexeme in a sentence, only a limited set of other lexemes and grammatical forms can be selected.[1]

The initial divide between the lexicon and the grammar is rooted in the Chomskyan view of the language architecture in which the only generative engine is syntax (the ruler of human language for its complexity and hypothesized universality), while the lexicon is only a list of items fed into the syntax according to certain principles.[2] It is worth noting that the divide between the lexicon and grammar, as a matter of principal, resembles the divide between the levels of linguistic analysis within the grammar, while there is no such partition within the conceptual structure.

Just about at the same time when Chomsky proposed the strict division between the lexicon and the grammar, Charles Fillmore advocated a different view emphasizing the crucial role of semantics in grammar. One of the founders of Construction Grammar, he formulated the theory of frame semantics[3] grounded in the idea that words are one kind of construction alongside others that come in all sizes. Ray Jackendoff, arguing for a decompositional theory of lexical or conceptual semantics, follows the lead of these early investigations. He elaborates that the conceptual structure, along with the phonological structure, is as generative as the syntactic one.[4]

In this article, as well as in several others, I adopt Jackendoff's Parallel Architecture (PA) (Jackendoff 1983, 1987, 1990, 1997, 2002) as a theoretical framework. The basic premise of this approach is that linguistic structure is determined by three independent generative systems – phonology, syntax, and semantics, and the links between them. This framework is particularly suitable for investigations of the relation between the lexicon and the grammar because it is grounded in the theory of Conceptual Semantics, an explicit psychologically based theory of meaning, as well as the theory of Simpler Syntax that Jackendoff developed with with Culicover (Culicover, Jackendoff 2005). Both tiers of this research are relevant for the present inquiry because they examine the implications of relations of meaning to linguistic form, while at the same time proposing an alternative to the mainstream generative grammar.

In Parallel Architecture, a word is seen as a small-scale interface principle that links pieces of phonological, syntactic, and semantic structure. The separate lexical interface is not anticipated. My take on this is to view the relationship between lexicon

1 See more in Schönefeld 2001, ch. 3 and 4.
2 See Chomsky 1957, 1980, 1995. For criticism of this view see Jackendoff 2006.
3 See Fillmore 1982; Kay, Fillmore 1999.
4 See Jackendoff 1983, 1987, 1990, 2002; Pinker 1989; Pustejovsky 1991.

and grammar within this framework as an umbrella for the complex and somewhat permeable multidimensional network of particular relations between chunks of lexicon and of grammar. The goal of this article is to underline the importance of disentangling the universal and language specific features that are relevant for an accurate description of lexical items. In order to exemplify my point of view, I present a small set of explanatory examples from Croatian and Slovene.

2 Why do we need grammar in the lexicon?

A dictionary based entirely on research into lexical and grammatical relations and grounded in a cognitive perspective[5] would require an extensive team to develop an elaborate model of the three-level description (phonological, morphosyntactic and semantic). As far as was possible in a one-volume dictionary produced by one person, the present Slovene-Croatian and Croatian-Slovene dictionary (Peti-Stantić 2014b) is established on these principles. Its main intention is to be a functional tool and a model for further investigations such as the one presented here.

By choosing examples from the class of nouns and verbs, I want to show that even the lexemes rich in semantic structures carry significant amount of grammatical information in their lexical form. I do not analyse grammatical parts of speech, nor idioms, which usually serve as prominent examples in discussions of the relation between grammar and lexicon.[6]

The analysis of the word CAT within the PA serves as a template for my analysis of Croatian and Slovene nouns. The initial assumption is that phonological and semantic content have inherent formation rules that do not belong to syntax. Rather, each component appears in its proper structure, linked to other two structures, which makes the account of combinatorial properties of lexemes explicit. The analysis of the noun CAT in English is simple, because its syntactic structure does not extend beyond N, so the simple linking mechanism between the three representations suffices for a description.

Phonology: /kæt/$_1$
Syntax: [N]$_1$
Semantics: [FELINE, PET, etc.]$_1$

I will show how the mismatches between the chunks on different structural levels cannot be detected without an explicit model of lexical analysis that accounts for all three representations. For that, I start off with the examples of transparent nouns in Slovene and in Croatian. Finding a straightforward one-to-one relationship between the phonology, morphosyntax and semantics within one linguistic system, particularly

5 See more on perspectives in Jackendoff 2012.

6 Just to give a flavour of types of idiomatic constructions that would play a role in a discussion of mismatches between the levels, compare the following: Slo *grem žvižgat rakom*; literal meaning: 'I will go whistle to the crabs'; actual meaning: 'I will fail'; Cro *ubit ću oko*; literal meaning: 'I will kill an eye' actual meaning: 'I will go to sleep.'

in a comparative context, is not an easy task. Here is one of the rare one-to-one examples in Croatian and Slovene.

Croatian	PREŠA
Phonology	/preša/$_1$
Morphosyntax	[$N_{fem\ inan}$]$_1$
Semantics	[STRAINER; SQUEEZER]$_1$
Slovene	STISKALNICA
Phonology	/stiskaːlnica/$_1$
Morphosyntax	[$N_{fem\ inan}$]$_1$
Semantics	[STRAINER; SQUEEZER]$_1$

Croatian PREŠA and Slovene STISKALNICA form a simple case of lexemes that match in their semantic and morphosyntactic description, although they differentiate entirely in their respective phonologies. This means that they mean the same or approximately the same, and have the same morphosyntactic features – both lexemes are nouns, both are feminine and both designate an inanimate object.[7] This is a typical case in comparing lexemes that belong to two language systems. Non-matching settings among two languages are conceivable at every level. For that, the answer to the question on the minimal requirements needed for comparison of lexemes is somewhat unexpected and most certainly language-specific. What we can compare depends on the entire system which, of course, develops based on the relation of its parts.

The next stage in comparing lexemes is an example with matching semantics and non-matching phonology (ORMAR vs. OMARA) and morphosyntax (masculine vs. feminine gender).

Cro	ORMAR
Ph	/ormaːr/$_1$
MS	[$N_{masc\ inan}$]$_1$
S	[CONTAINER FOR SMALL OBJECTS, SUCH AS CLOTHES AND THE LIKE]$_1$
Slo	OMARA
Ph	/ɔmaːra/$_1$
MS	[$N_{fem\ inan}$]$_1$
S	CONTAINER FOR SMALL OBJECTS, SUCH AS CLOTHES AND THE LIKE$_1$

The relation between the examples that follow is even more complicated. They match in phonology and in morphosyntax, but diverge in semantics in one language, and then diverge at all levels but semantics in Slovene.

7 Although the feature »inanimate« is not necessary for the feminine and neuter in Croatian and Slovene as it is in masculine, for morphosyntactic reasons I employ it in all three cases for the sake of systematicity. Also, since the contrastive analysis was never performed within PA, this is tentative attempt and the indices would probably be different under more meticulous account.

Cro	PISMO	PISMO
Ph	/pi:smo/$_{1,2}$	/pi:smo/$_{1,2}$
MS	[$N_{neut\ inan}$]$_{1,2}$	[$N_{neut\ inan}$]$_{1,2}$
S	[LETTER]$_1$	[SCRIPT]$_2$
Slo	PISMO	PISAVA
Ph	/piːsmɔ/$_1$	/pisaːva/$_2$
MS	[$N_{neut\ inan}$]$_1$	[$N_{fem\ inan}$]$_2$
S	[LETTER]$_1$	[SCRIPT]$_2$

As is clear from the indices that designate the relations, in Croatian there is one lexeme with one phonological representation (indexed with 1,2) that has one morphosyntactic representation (again indexed with 1,2), which splits at the level of semantic representation. In Slovene we start off with two lexemes for the same semantics and progress in the same way through all the levels. What makes this even more appealing between languages is the fact that Croatian PISMO with semantics indexed 2 and Slovene PISAVA indexed the same way differ in that one is neuter and the other one is feminine. There is no way of knowing this without making it explicit.

Before discussing the relations within the verbal system, I once again wish to stress the fact that the set of necessary categorical relations, both at the morphosyntactic and semantic levels, is language-specific and depends on the broadly understood agreement potential within every individual language.[8] Beyond the most general description of lexemes as belonging to certain parts of speech, there are many categories that play an important role in the morphosyntax of some languages, while not playing a role in others. The aforementioned is an example of gender, which plays a significant role in agreement in Croatian and Slovene, but does not play a role in English. Accordingly, there is no need to put this information in the description of a lexeme in English, but it is necessary to put it in the description of Croatian and Slovene. Another stock in trade is animacy. While some treat it as a semantic feature and the others as a syntactic one, many see animacy as a principle in between morphosyntax and semantics. »The typical way in which animacy shows up in Slavic languages is giving the accusative case of animate nouns the form of the genitive.« (Fillmore 1980: 43) The separation of the semantic and syntactic features of animacy and the decision about which ones are relevant for a specific language should lead to a more accurate description and, consequently, a more precise network of the agreement potential for lexemes.

The following description of verbs is based on Jackendoff's analysis of the verb DEVOUR. This analysis is not that simple, because, unlike nouns, the syntactic structure of verbs extends beyond V into the VP, which needs to be explicitly stated.[9]

8 See more on that in Haspelmath (2007, 2014).

9 A notational commentary prior to analyzing Slavic examples: as one can see, at the morphosyntactic level we are explicit about the fact that for the description of verbs one needs an extension beyond V to VP. Underlining NP signals obligatory argument.

Ph	/devawr/$_1$
MS	$[_{VP}\ [V]_1\ \underline{NP}_3]_1$
S	[DEVOUR (Agent: X_2, Patient: Y_3)]$_1$

The following verbs have the same or almost the same phonological representation in both languages, but differ at one or more of other levels.

Slo	POMILOVATI[10]
Ph	/pɔmilɔva:ti/$_1$
MS	$[_{VP}\ [V_{imperf}]_1\ \underline{NP,}\ Acc_3]_1$
S	[FEEL SORRY$_1$ (Agent: X_2, Patient: Y_3)]$_1$

Cro	POMILOVATI[11]
Ph	/pomilovati/$_{2,3}$
MS	$[_{VP}\ [V_{perf}]_{2,3}\ \underline{NP},\ Acc_4]_{2,3}$
S	[CARES$_2$; PAT$_2$ (Agent: $\underline{X}_2$, Patient: Y_4)]$_2$

Cro	POMILOVATI
Ph	/pomilovati/$_{2,3}$
MS	$[_{VP}\ [V_{perf}]_{2,3}\ \underline{NP},\ Acc_5]$
S	[GRANT AN AMNESTY$_3$ (Agent: $\underline{X}_2$, Patient: Y_5)]$_3$

When comparing Slovene and Croatian verbs, we notice that the Slovene verb is imperfective, while the Croatian one is perfective. Without the morphosyntactic information at the lexical level, this would not be visible. Also, although all three verbs have the same root, their meanings departed from contingent shared meaning so that they should be defined separately. The indices allow us to explicitly link respective representations to each other.

The following final example is the most instructive so far, because it enables us to demonstrate the multidimensional relation between verbs that share certain – but not all – chunks of phonology, morphosyntax and semantics. In Croatian, there are two verbs PASTI[12] that constitute a phonological minimal pair. These verbs have the same repertoire of phonemes, but differ in what we can call suprasegmental features (accent) and/or quantity of only one phoneme. So, consequently, the difference is only in one segment's length at the phonological level. Although this might seem a very rare situation, it is, in fact, much more common than we are usually aware, even across languages. This is the case because it is often blurred through the symbolic repertoires that do not enable us to reach even the segmental level, let alone the suprasegmental one.

Cro	PASTI	PASTI
Ph	/pasti/$_1$	/pa:sti/$_2$
MS	$[_{VP}\ [V_{perf\ intrans}]_1]_1$	$[_{VP}\ [V_{perf}]_2\ NP_{inan},\ Acc_3]_2$
S	FALL$_1$ (Agent: X_2)	GRAZE$_2$ (Agent: X_2, Patient: Y_3)

10 Dictionary account for Slo-Cro pomilovati = sažalijevati, žaliti.

11 Dictionary account for Cro-Slo pomilovati = 1. pobožati, pogladiti, podragati 2. amnestirati, pomilostiti.

12 Sentences would be, for the first one: *Dijete je palo*. 'The child has fallen.' For the second one: *Krave pasu (travu)*. 'The cows graze (the grass).' Notice that the argument is not obligatory with the second verb.

Slovene has both of these verbs, but with somewhat different distribution of meaning and with an interesting reflexive addition to the second one

Slo	PASTI
Ph	/paːsti/$_1$
MS	[$_{VP}$ [$V_{perf\ intrans}$]$_1$]
S	FALL$_1$ (Agent: X_2)

Slo	PASTI	PASTI SE
Ph	/paːsti/$_3$	/paːsti/ /sɛ/$_4$
MS	[$_{VP}$ [V_{perf}]$_3$ $\underline{NP}_{anim\ non\text{-}human}$, Acc_3]$_3$	[$_{VP}$ [$V_{perf\ reflex}$]$_1$][13]
S	[CAUSE [X](GRAZE$_3$ (Agent: X_2, Patient: Y_3)	[GRAZE$_4$ (Agent: X_2)]

There are many points of interest here. I just want to underline that, if we did not perform such a detailed morphosyntactic analysis, we would not be able to compare the Croatian and Slovene verbs PASTI. We would also not be able to see that the argument in Croatian is not obligatory but if it occurs, it is inanimate and in Slovene it is obligatory, animate and non-human.[14] This analysis, of course, could go further. Suffice to say here, that examples from the Gigafida online corpus of Slovene show that the uses of reflexive and non-reflexive variant of the verb with the same meaning are mutually exclusive. Translation to Croatian furthermore shows that in contrast to Slovene, we need to introduce a new verb in the Croatian network, namely NAPASATI.[15] Although the reflexive option is generally available in Croatian, it is not actualized with this verb.

Although I cannot show here the extent of the necessary inclusion of grammatical information in the lexicon, I hope that these examples already gave a flavour of what

13 Although many verbs in Slovene can be both transitive and intransitive depending on their use in a sentence, all reflexive verbs are intransitive. Therefore it is redundant to mark the category of transitivity in this case.

14 For that reason, if we wanted to say a simple Croatian sentence *Krave pasu travu* in Slovene, this sentence would not be, as we might think without such a detailed analysis, **Krave pasejo travo*, than *Krave se pasejo na travi*, where *na trav*i is not an argument, but adverbial PP.

15
1.1 Slo *Ingo Robič pase svoje tri krave.*
1.2 Cro **Ingo Robič pase svoje tri krave.*
1.3 Cro *Ingo Robič napasa/vodi na pašu svoje tri krave.*
'Ingo Robič is grazing his three cows.'
2.1 Slo *Marjanca je pasla krave na travniku.*
2.2 Cro **Marjanca je pasla krave na livadi.*
2.3 Cro *Marjanca je napasala krave na livadi.*
'Marjanca was grazing the cows in the meadow.'
3.1 Slo *Ob hišah se pasejo ovce in krave.*
3.2 Cro **Pored kuća se pasu ovce i krave.*
3.3 Cro *Pored kuća pasu ovce i krave.*
'The sheep and cows were grazing near the houses.'
4.1 Slo *Na travniku se pasejo koze in krav*e.
4.2 Cro **Na livadi se pasu koze i krave.*
4.3 Cro *Na livadi pasu koze i krave.*
'The goats and cows were grazing in the meadow.'

else might be of interest at the intersection of the lexicon and the grammar in Croatian and Slovene.

3 Conclusion

The amount and the range of grammatical information, primarily morphosyntactic, that should be included in every scholarly description of a lexeme greatly exceeds what most linguists expect. The characterization of the lexicon in terms of its functional properties remains incomplete and inconsistent without them and such information can only serve as a shallow description for those who nonetheless intuitively know how to use lexemes. Surely this should not be the goal of an accurate account. The extent and the depth of grammatical information needed for rigorous description of the lexical inventory of the language in question depend primarily on the structure of a specific language, and then, to some extent, on the task. When speaking of specialized tasks, I primarily have in mind cross-linguistic contrastive studies investigating sets of constraints for the universality and language specificity of any given language. For that reason, in order to carry out detailed research into the range of grammatical information needed for the meticulous lexical description of Slovene (as well as Croatian, or any other language), it is necessary to dig deeper into the grammar of the specific language. It is also necessary to investigate the relations of this language with closely related ones and then, at the next level, more distantly related languages.

Bibliography

BATES, Elizabeth, GOODMAN, Judith C., 1997: On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. *Language and Cognitive Processes* 12/5–6. 507–584.

BRESNAN, Joan, 1982: Control and Complementation. Joan Bresnan (ed.): *The Mental Representation of Grammatical Relations*. Cambridge: The MIT Press. 282–390.

BRESNAN, Joan, 2001: *Lexical Functional Syntax*. Blackwell.

CHOMSKY, Noam, 1957: *Syntactic structures.* The Hague: Mouton.

CHOMSKY, Noam, 1980: *Rules and representations.* New York: Columbia University Press.

CHOMSKY, Noam, 1995: *The minimalist program.* Cambridge: MIT Press.

CULICOVER, Peter W., JACKENDOFF, Ray 2005: *Simpler Syntax.* Oxford: Oxford University Press.

FILLMORE, Greville G., 1980: Animacy in Russian and in other Slavonic languages: Where syntax and semantics fail to match. Ch. V Chvany, R. D. Brecht (eds.): *Morphosyntax in Slavic.* Bloomington: Slavica Publishers.

FILLMORE, Charles J., 1982: Frame semantics. *Linguistics in the morning calm*. The Linguistic Society of Korea. Seoul: Hanshin. 111–137.

FILLMORE, Charles J., KAY, Paul, O'CONNOR, Cathy, 1988: »Regularity and idiomaticity in grammatical constructions: The Case of Let Alone.« *Language* 64/3. 501–538.

Gigafida: http://www.gigafida.net/

GOLDBERG, Adele, 2006: *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.

HASPELMATH, Martin, 2007: Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11/1. 119–132.

HASPELMATH, Martin, 2014: The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. Michael Tomasello (ed.): *The New Psychology of language Structure*: *Cognitive and Functional Approaches to Language Structure.* Psychology Press and Routledge Classic Editions. 211–242.

JACKENDOFF, Ray, 1983: *Semantics and Cognition*. Cambridge: MIT Press.

JACKENDOFF, Ray, 1987: *Consciousness and the Computational Mind*. Cambridge: MIT Press.

JACKENDOFF, Ray, 1990: *Semantic Structures*. Cambridge: MIT Press.

JACKENDOFF, Ray, 1992: *Languages of the Mind: Essays on Mental Representation*. Cambridge: MIT Press.

JACKENDOFF, Ray, 1993: *Patterns in the Mind: Language and Human Nature*. New York: Harvester Wheatsheaf.

JACKENDOFF, Ray, 1997: *The Architecture of the Language Faculty.* Cambridge: MIT Press.

JACKENDOFF, Ray, 2002: *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

JACKENDOFF, Ray, 2010: *Meaning and the Lexicon: The Parallel Architecture 1975–2010*. Oxford: Oxford University Press.

JACKENDOFF, Ray, 2011: Alternative Minimalist Visions of Language. R. D. Borsley, K. Börjars (eds.): *Non-Transformational Syntax: Formal and Explicit Models of Grammar.* Wiley-Blackwell, Oxford. 268–296.

JACKENDOFF, Ray, 2012: *A User's Guide to Thought and Meaning*. Oxford: Oxford University Press.

KAY, Paul, FILLMORE, Charles J., 1999: Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* 75/1. 1–33.

LAMB, Sydney M, 1966: *Outline of Stratificational Grammar.* Washington DC: Georgetown University Press.

PETI-STANTIĆ, Anita, 2014a: Pravi in lažni slovenski in hrvaški prijatelji. Hotimir Tivadar (ur.): *Prihodnost v slovenskem jeziku, literaturi in kulturi*. 50. seminar slovenskega jezika, literature in kulture. 201–209.

PETI-STANTIĆ, Anita, 2014b: *Veliki suvremeni slovensko-hrvatski i hrvatsko slovenski rječnik*. Zagreb: Mozaik knjiga.

PINKER, Steven, 1989: *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge: MIT Press.

PUSTEJOVSKY, James, 1991: The Generative Lexicon. *Computational Linguistics* 17/4. 409–441.

SADOCK, Jerrold M., 1991: *A Theory of Parallel Grammatical Representations*. Studies in Contemporary Linguistics. The University of Chicago Press Books.

SCHÖNEFELD, Doris, 2001: *Where lexicon and syntax meet.* Trends in Linguistics. Studies and Monographs 135. Mouton de Gruyter.

van VALIN, Robert D., Jr. (ed.), 1993: *Advances in Role and Reference Grammar.* Amsterdam: Benjamins.

van VALIN, Robert D., Jr., 2003: *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

van VALIN, Robert D., Jr., la POLLA, Randy, 1997: *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.