Kirill Vasiltsov
Kyoto University
vasiltcov.kirill.36z(at)kyoto-u.ac.jp

Labeling in narrow syntax and sensory perception: reasons and consequences

In this paper I explore the nature of labeling. First, I argue against approaching the problem of labeling with minimal search algorithms and suggest that in fact lexical items are fully responsible for label determination. I show how label affects the workings of narrow syntax by focusing on unusual agreement phenomena in Russian. Next I investigate the status of labeling within our mind by focusing on examples from temporal, visual and musical cognition. I argue that the purpose of labeling (in a broad sense) is establishment of asymmetry, which allows us to make perceptual/meaningful distinctions about the world and to make objects on which a cognitive faculty operates usable in other domains. It is in this sense that label is not only "conceptually necessary" but also empirically necessary for objects generated by narrow syntax, which have to be interpreted at the interfaces that relate language to the rest of the cognitive faculties. Throughout the paper I mostly adopt Hornstein's (2009) view of labeling.

1. Introduction

In earlier frameworks phrase structure was built in accordance with rewrite rules (or by X-bar-theoretic projections), such as VP → V N, which "automatically" provided a label for a generated structure. In MP, for a number of reasons phrase structure rules were abandoned in favor of Merge - an operation that forms larger units out of those already constructed. Applied to α and β, Merge yields a new object γ={α, β}. It is obvious, however, that e.g. verb phrases and noun phrases are interpreted and behave very differently, so something must also indicate the relevant properties of the object constructed. Chomsky (1994) assumes that Merge yields at least γ={H, {α, β}} where H is the label, identical to the head that "projects" (either α or β). The label H determines γ's syntactic behavior and the way γ is interpreted at the interfaces.

Whether labeling is a distinct fundamental operation or not is an open question. Hornstein (2009) suggests that Label is a distinct operation, while Chomsky (2008, 2013) suggests that labels are identified by what he calls "minimal search". Chomsky presents a Labeling Algorithm (1):

(1) a. In {H, α}, H an LI, H is the label
b. If α is internally merged to β, forming {α, β}, then the label of β is the label of {α, β}

While (1) succeeds to a certain extent in labeling structures without positing Label as

a distinct operation or making Merge itself determine the label, the algorithm is flawed for a number of reasons. First of all there is *no way* to determine what a head/LI is in bare phrase structure, because the difference between zero-level projections and maximal projections is stipulated in X-bar terms (projection etc.). Second, (1b) implies that internal Merge and external Merge are different operations, which they are likely not - the point repeatedly emphasized by Chomsky himself. Apart from the flaws above, (1) is also unable to label {XP,YP} structures. Chomsky (2013) suggests that "feature sharing" takes place between e.g. a subject and TP and as a result the structure is labeled as <φ, φ>. But why should this instance of labeling be so different from others? Moreover, it again implies that XPs are fundamentally different from Xs which is a somewhat X-bar theoretic stipulation. It is also unclear how exactly feature sharing between the two contributes to semantic (or phonological) interpretation. Finally, this labeling algorithm would fail to label the "lowest" constituent in the structure, which is not of the form {H,XP}(for example {V,N}). If we dispense with X-bar theory then the theoretical status of the algorithm becomes unclear.

It becomes clear that labeling is not an easy task given the minimal theoretical distinctions available in MP. Yet we face the fact that labeling is crucial for semantic (and possibly phonological) interpretation. It is very possible that interfaces need the merged objects to be labeled in accordance with SMT, which makes labels at least "conceptually necessary". In the last section I will show that something akin to labeling can be found in other cognitive faculties and conjecture that labeling not only serves to make narrow-syntactic objects "usable" at the interfaces, but also make other objects of mind usable elsewhere. This is not to say that labels cannot have any effect on what happens in narrow syntax and below I will show that it indeed does.

One obvious difficulty that arises if we abandon X-bar theory is that most A and A'-movement is movement of *phrases* such as DP. With Merge, the simplest "DP" that can be generated looks like {the, cat}. Suppose that either 'the' or 'cat' has relevant features to be checked in Spec/T or Spec/v. Whatever theory of movement one has in mind - free Merge or Merge triggered by probes - it looks as if *whole phrases* and not actual feature bearing items are "targeted". In other words, among convergent derivations we never see structures where either 'the' or 'cat' successfully checks the relevant features, while leaving the other in its initial position. Why should it be this way if both items are equally accessible by a probe? One possible explanation is that merged objects are labeled immediately and that label not only indicates the relevant properties of the generated object but is also responsible for syntactic operations such as feature checking/valuation. Moreover, if we adopt Hornstein's (2009) view of labeling, then label is also what enables recursive structure building and therefore syntax needs objects to be labeled immediately. If labeling is indeed what makes the product of Merge atomic, then it explains why we mostly see movement of whole *phrases*. This is much closer to the earlier views of Chomsky (1995)[1] or Chomsky (1998).

Unless proven otherwise, I assume that lexical items themselves are *fully* responsible for label determination and that the details of labeling are rather obscure and cannot be dealt with by simple algorithms like the one above. Instead of finding ways to refine Chomsky's algorithm, I suggest that given the minimal theoretical distinctions we have

[1] Strictly speaking, the sub-label, as a feature bundle, was involved in checking operations.

today, *any* labeling algorithm by minimal search is impossible and that constituents are labeled immediately. Furthermore, labels have narrow-syntactic reality and it can be seen by looking at some interesting agreement phenomena in Russian.

2. Agreement failure

In Russian, subjects and verbs agree in gender, person and number. It is true of all kinds of verbs - transitive, unergative and unaccusative. Sometimes, however, "unmarked" singular neuter agreement appears on verbs. This seems to happen only when the subject is quantified. In literature, such unmarked agreement is taken to be an instance of non-agreement (Pesetsky 1982, Neidle 1988, Franks 1995, Boskovic 2006, Pereltsvaig 2006). Among quantifiers that appear in subjects are numerals greater than *pjat'* (five)[2] and approximate quantifiers such as *mnogo* (many) / *malo* (little/few) / *neskol'ko* (few) / *nemerjano* (very many) / *dostatochno* (enough). Note that agreeing and non-agreeing subjects are not in complementary distribution - quantified subjects *may* agree with a verb or not. Closer investigation reveals that this variation is not due to mere optionality.

2.1. Previous approaches to QP/NP distinction

The basic (non-)agreement pattern is presented in (2):

(2) *Student* *prish-ol* *na zanjatije*.
Student.SG.M came-SG.M to class
'A student came to class.'

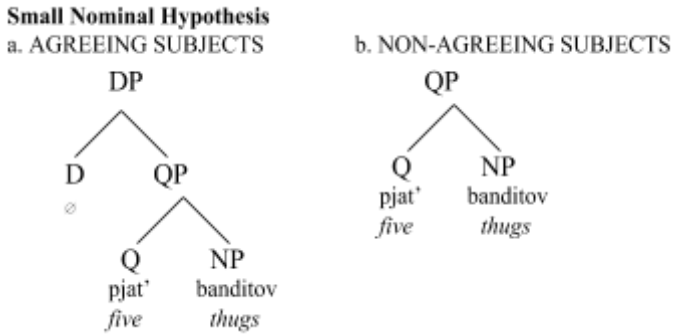
Mnogo studentov *prishl-i / prishl-o* *na zanjatije*.
Many student.PL came-PL / came-SG.NEUT to class
'Many students came to class.'

One explanation for the observed pattern is given in Pesetsky (1982). He claims that when subjects agree with the verb they are NPs and when they don't agree they are QPs. Pesetsky motivates it by considering how ECP and selection affect quantifier raising. The basic idea is that unlike NPs, QP subjects must undergo QR and after raising its trace becomes NP to satisfy selectional requirements. Since the trace of QP subject is not lexically governed, the only remaining option provided by ECP is antecedent-government. However due to "category mismatch" antecedent-government is impossible too. As a result QP subjects of transitive verbs cannot undergo QR, while QP subjects of unaccusative verbs can, because they originate in a lexically governed position. Pesetsky uses this reasoning to explain several scope phenomena in Russian. The problem with his analysis is that his acceptability judgements are *wrong*, as pointed out in Pereltsvaig (2006). This and the fact that government is no longer considered to be a fundamental relation render his analysis unsatisfactory. Pesetsky's analysis was

[2] I ignore numerals lesser than five, because they regularly induce paucal agreement, that is not relevant to our investigation here.

adopted by Franks (1995) where he developed the idea that agreeing NP subjects raise to Spec/TP while non-agreeing QPs stay in Spec/VP. However, simple observation of word order in Russian dictates otherwise: quantified subjects can and do occur in front of modal verbs which are undoubtedly T[3]. Pereltsvaig (2006) comes to a conclusion that there is not enough evidence that non-agreeing QPs stay in Spec/VP and argues that both kinds of subjects in fact occupy the same position. She also abandons the NP/QP distinction and proposes instead that agreeing quantified subjects have a phonologically null D above them, whereas non-agreeing QPs (small nominals) lack it as shown in (3). D is believed to value φ-features of N which enters a derivation unvalued. Without D, QP is unable to agree. Pereltsvaig motivates it by assuming that DPs are referential and small nominals are not, since only a fully valued set of φ-features makes something refential. The non-referentiality is reflected in impossibility of small nominals to bind reflexives and to control PRO.

(3)



[from Pereltsvaig (2006)]

While some findings about QPs and control (extensively investigated by Franks & Hornstein (1992)) are indeed very interesting, a closer look at the simplest instances of control reveals that so-called non-referential subjects in fact *can* control PRO as in (4).

(4) [*Pjat' devushek hotel-o* [PRO *kupit' etu jubku*]].
Five girls.PL wanted-SG.NEUT buy.INF this skirt
'Five girls wanted to buy this skirt.'

Small nominals can also bind reflexives, contrary to what is proposed in Pereltsvaig (2006). But more importantly, I want to focus on her use of the notion of "referentiality". Does "referential" mean having semantic content or picking up an object of the real world? As Chomsky constantly argued, lexical items do *not* refer to objects of the world - they are only *used to refer* to them. In this sense, both QPs and DPs are non-referential. But this is probably not what Pereltsvaig intended. For her, it

[3] The possible explanation for why QPs occur in front of modal verbs when they are expected to be in Spec/VP is that only phonological matrix (as a bundle of phonological features) of QP is copied to Spec/TP for some unknown reasons. However while Q and N have their own phonological matrices stored in lexicon, there does not exist a matrix for the whole QP. That is, there is nothing to target if it is indeed only phonological material that is targeted.

seems, what matters is having (or not having) an *extension* in classical sense. I take it then that from her point of view small nominals do not refer to / do not pick up anything *certain* (any token) in the real world. However consider the sentences below:

(5) a. *Pjat' etix studentov prishl-o / prishl-i na zanjatije*.
Five these students.PL came.SG.NEUT / came.PL to class
'These five students came to class.'
b. *Eti pjat' studentov * prishl-o / prishl-i na zanjatije*.
These five students.PL came.SG.NEUT / came.PL to class
'These five students came to class.'

(6) *Skoljko studentov prishl-i / prishl-o na zanjatie?*
How many students came.PL came-SG.NEUT to class
'How many students came to class?'

In (5a) there was a past event of coming of *these* students to class. This entails that there were *certain* students who came to a *certain* class. Yet non-agreement is possible even in this sentence. I do not see how non-agreeing 'mnogo etix studentov' can be non-referential at all in Pereltsvaig's sense. Moreover, the wh-phrase in (6) by virtue of being a question and by definition is clearly not D(efinite), yet it can φ-agree with the verb. These simple cases cannot be accounted for by the hypothesis in (3) and suggest that other factors are at work. Agreement and non-agreement certainly affect semantic interpretation (such as group/individuated interpretation differences) but both agreeing QPs and non-agreeing QPs have *some* semantic content and can be used to refer while not picking up any entity in the real world. Still, (5b) indicates that D is somehow responsible for agreement at least in some cases. In general, it seems that QP/NP (or QP/DP for Franks(1995)) distinction should be maintained but for reasons not previously considered in the literature.

2.2. Label as agreement blocker in Russian

There is already a plenty of evidence in literature that QP/NP distinction is real in Russian. Below I provide new evidence that the distinction is caused solely by the choice of label for {Q,N}. The choice happens to be random, but influences agreement depending on what exactly is chosen. Once Q is a label, it somehow hides N from and its φ-features from "outside", blocking agreement. Let us investigate each case closely.

2.2.1. Quantifier 'kucha'

'Kucha' (a bunch/heap/pile) can be used either literally (in which case it is a feminine noun) or metaphorically as a Q head. Due to homophony we can better understand how categories influence agreement. Whenever 'kucha' is used literally, agreement never fails and a verb agrees with it. When it is used as a Q head, feminine agreement is unacceptable. This is shown in (7) below.

(7) a. *Kucha studentov prishl-o / *?prishl-a na meroprijatije.*

Bunch student.PL came-SG.NEUT / came-SG.FEM to event
'A bunch of students came to the event.'

b. *Kucha kamnei *lezhal-o / lezhal-a / lezhal-i na skamejke.*
Heap.FEM stone.PL lied-SG.NEUT / lied-SG.FEM / lied-PL on the bench
'A heap of stones was lying on the table.'

The unacceptability of feminine agreement in (7a) can be explained by safely assuming that unlike nouns, Q does not have any φ-features that can enter into agreement operations. In (7b), on the other hand, we are dealing with a subject that consists of two nouns. Since both nouns apparently have φ-features then verb agreement will depend on which one is the label and non-agreement is impossible. This is exactly what we observe in (7b). Therefore it is reasonable to believe that the Q label in (7a) might block agreement.

2.2.2. Quantifier scope

(8a) below shows that 'ne' ("not") cannot scope over 'kucha'. This is to be expected if 'kucha' is used literally, and the subject is unquantified. Then, along the lines of Fox (2000), we can say that raising of 'ne' in (8a) results in vacuous quantification, and is prevented by Scope Economy. This is not the case in (8b), where scope variation is allowed. The variation suggests that the subject of (8b) is quantified and must be a QP. The singular neuter agreement in this sentence in turn supports the idea that Q blocks agreement between φ-features of the noun and the verb.

(8) a. *Kucha kamnei ne lezhala / *lezhalo na skamejke.*
Heap.FEM stones.GEN NEG lied.SG.FEM / lied.SG.NEUT on the bench
'A heap of stones was not lying on the table.' (*not > heap, heap > not)

b. *Mnogo studentov ne prishl-o na zanjatije.*
Many students NEG came.SG.NEUT to class
'Many students did not come to class.' (?not > many, many > not)

2.2.3. Secondary predication and Case

In Russian, NPs agree with secondary predicates in gender, number and case. Secondary predicates also have a "default" form in which case they are inflected with instrumental case. When agreement is impossible the default form must be chosen. Cases of non-agreement were investigated by Franks & Hornstein (1992) and Franks (1995). However they mostly focused on control environments and PRO. One case not yet attested is the non-agreement between quantified subjects and secondary predicates. Consider (9):

(9) a. *Neskol'ko studentov prishl-i na meroprijatije pjanymi / pjanyje.*
Few students.NOM came.PL to event drunk.INS / drunk.NOM
'Few studens came to the event drunk.'

b. *Ja zastal studentov pjanymi / pjanyx / *pjanyje.*
I caught students.ACC drunk.INS / drunk.ACC / drunk.NOM
'I caught students drunk.'

c. *Neskol'ko studentov prishl-o na meroprijatije pjanymi / *pjanyje.*
Few students came.SG.NEUT to event drunk.INS / drunk.NOM
'Few studens came to the event drunk.'

In both (9a) and (9b) both default form and agreeing form are allowed. When secondary predicate agrees with the subject it takes nominative form and when it agrees with the object it takes accusative form. On the other hand, (9c) does not allow any form except the default one. This indicates that the subject does not bear *any* case. The failure of case agreement between e.g. a subject QP and T is to be expected if case features are checked along with φ-features as suggested by Chomsky (1995). Next consider (10):

(10) a. *Mnogo ljudej obmanyva-jut sami sebya.*
Many people deceive.PL selves.PL.NOM REFL.ACC
'Many people deceive themselves.'

b. **Mnogo ljudej obmanyva-jet sami sebya.*
Many people deceive.SG.NEUT selves.PL.NOM REFL.ACC
'Many people deceive themselves.'

c. **Mnogo ljudej obmanyva-jet samo sebya.*
Many people deceive.SG.NEUT self.SG.NEUT.NOM REFL.ACC
'Many people deceive themselves.'

d. *Okno zakryl-o samo sebya.*
Window.SG.NEUT closed-SG.NEUT self.SG.NEUT.NOM REFL.ACC
'The window closed itself.'

Russian 'sam-' can be used to emphasize the reflexive reading. When it is used it agrees in φ-features and case with the relevant NP just like secondary predicates. Whether 'sam-' really is a secondary predicate or not is a tricky question. One property of 'sam-' that is especially interesting is that unlike other secondary predicates 'sam-' does not have a default form. Then we can predict that 'sam-' simply cannot occur in the sentence if the subject does not have any case. This is exactly what we see in examples above. In (10a) the subject successfully agrees with the verb and with 'sam-' in plural number and nominative case. In (10b), although the QP is the subject it fails agreement with the verb and 'sam-' cannot bear nominative case. In fact any form of 'sam-' is impossible in (10b) just as predicted. It might seem that such unacceptability is due to the fact that the verb is singular neuter, so 'sam-' must also take this form. (10c), however, shows that this is not the case. Moreover, as (10d) indicates, 'sam-' does agree with unquantified sg-neut subjects that normally agree with the verb. This allows us to

exclude the possibility that sg-neut agreement on verbs is due to idiosyncratic φ-features of Q. Above sentences provide more evidence that non-agreeing QP subjects lack any Case. Hence, if Q indeed blocks agreement then it is reasonable to believe that it blocks *any* kind of agreement.

2.2.4. Relativization

When NP is relativized, a relative pronoun φ-agrees with the NP and bears the case that the NP had prior to relativization. Just in case there is no subject-verb (relative pronoun-verb) agreement the subject cannot be relativized. Only agreeing NPs can be relativized as shown in (11).

(11) a. *Mnogo ljudei, kotorye kupil-i igru.*
Many people who.PL.NOM bought.PL game
'Many people who bought the game.'

b. *Mnogo ljudei, kotoryx ja uvidel v metro.*
Many people, who.PL.ACC I saw in subway
'Many people whom I saw in subway.'

c. *Mnogo ljudei, kotorym ja dal deneg.*
Many people, whom.PL.DAT I gave money
'Many people whom I gave money'.

d. **Mnogo ljudei, kotorye kupil-o igru.*
Many people who.PL.NOM bought.SG.NEUT the game
'Many people who bought the game.'

From (11d), where the relative pronoun is in nominative case on analogy to (11a), it is clear that non-agreeing QPs cannot be relativized. This supports the assumption that non-agreeing QPs, despite being subjects, do not bear any Case. One more interesting observation is made in (12):

(12) *[*Mnogo ljudei, kotorye kupil-i igru,*] *prishlo na meroprijatie.*
Many people.PL who.PL.NOM bought.PL game came to event
'Many people who bought the game came to the event.'

Note that the internal structure of the subject is perfectly fine as it is identical to (11a). Yet the whole sentence is unacceptable. This is a very strong evidence for the idea that agreement is not simply optional. If, as shown above, the subject of (12) is NP, then it follows that N *must* agree when it is a label. The NP in (12) does not agree with the verb and renders the sentence unacceptable.

Let us finally return to sentences in (5) repeated here for convenience:

(5) a. *Pjat' etix studentov prishl-o / prishl-i na zanjatije*.
Five these students.PL came.SG.NEUT / came.PL to class

'These five students came to class.'

b. *Eti pjat' studentov* ***prishl-o** / *prishl-i* *na zanjatije*.*
These five students.PL came.SG.NEUT / came.PL to class
'These five students came to class.'

We can account for the difference between (5a) and (5b) by first making a simple assumption that word order within the subject reflects merge order in this case. Then in (5a) we have a structure {Q, {D, N}} and in (5b) we have {D, {Q, N}}. If, as I suggested above, labels – whatever their nature is – are responsible for feature checking/valuation, then we expect N, as a "φ-feature bearer", to agree with a verb when it is a label[4]. Next, we need few more assumptions to explain the data.

(13) 1. {Q,D} cannot be labeled and interpreted and thus represents an illegitimate object
2. {Q,N} can be labeled either as Q or N
3. {D,N} is labeled as N

From (13.1) it follows that merging D with {Q,N} will not yield a legitimate object unless the latter is labeled as N by (13.2). {D,N} is further labeled as N by (13.3), and since N bears all the necessary φ-features, structures as the subject in (5b) never fail agreement. On the other hand, in (5a) Q merges with {D,N} that is labeled as N and the resulting structure can be labeled either as Q or N by (13.2). Whenever it is Q, the subject will not agree with the verb. This explains why we observe variation only in (5a)-like cases. It looks as if once Q is a label, it somehow hides N and its φ-features from outside, blocking agreement.

The additional assumptions in (13) above are not trivial and appear somewhat stipulated. However it is possible to look for reasons why it should be this way. As for (13.1), one can think of Q and D as fundamentally of the same kind. DPs in Norwegian can have an inverse scope, as if they are quantified. Then the impossibility to label {Q,D} could follow from whatever does not allow e.g. {Q,Q}, {D,D}, or {C,C} structures. As for (13.2), I think that such "instability" can be observed elsewhere. Take "flying planes" as an example. The structure is {V,N} in both cases, but the string can be interpreted either as a verb phrase or as a noun phrase depending on whether the label is V or N. The "unstable" nature of {Q,N} can also be one possible cause of language variation. In Norwegian, for example, "bare" nominals never agree with the verb, whereas DPs (which have overt determiners) do[5]. One might think of bare nominals as QPs with a null quantifier - the idea present in the literature on nominals. Then we can say that Norwegian idiosyncratically "forces" a Q label in {Q, N} configurations, that blocks agreement with a verb. Russian, on the other hand, allows both options. This dimension of language variation seems more reasonable and lexically motivated than rigid narrow-syntactic constraints such as "nominals must merge with (an agreeing null) D in Russian, but do not have to do so in Norwegian".

[4] The stronger claim that X is able to agree *iff* X is a label will be formulated below.
[5] See Pereltsvaig (2006) for relevant data.

2.3. Other consequences of labeling

Agreement failure can be detected even in languages which lack rich morphology. Consider the following English sentences.

(14) Flying planes can be dangerous.
(15) Flying planes **are** dangerous.
(16) Flying planes **is** dangerous.

(14) is a famous ambiguous sentence, where interpretation depends on whether 'flying planes' is VP or NP. However, when only 'be' is used, the sentence can only be interpreted one way or another depending on how 'be' agrees. If ambiguity of (14) is just a matter of choosing a label for a simple set {V,N} then the fact that (15) and (16) are unambiguous suggests that the agreement pattern indicates the choice of label. We cannot interpret (15) as talking about an act of flying, and we cannot interpret (16) as talking about planes themselves. Then, whenever we talk about planes the subject seems to be NP (labeled as 'planes' in BPS terms) and agree in plural number with 'be'. Whenever it is VP, agreement between 'be' and 'planes' fails. This again suggests that label of VP, 'flying', blocks agreement of 'planes' with outside LIs. Moreover, whenever the subject is NP it *must* agree as we conjectured before, so that optional agreement for (15) is impossible. Note that this can also be used as evidence that agreement is structure-dependent.

A similar example can be found in Russian (17).

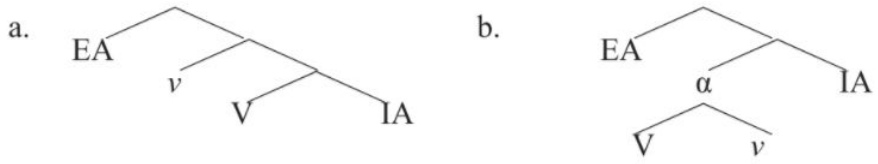
(17) [Sazhat' samolëty] byl-o opasno.
Land.INF plane.PL.ACC was.SG.NEUT dangerous
'Landing planes was dangerous'.

'Land' here is infinitive unlike English, but the logic is the same as in the examples above. 'Bylo' ('was') does not agree with plural 'samolëty' and this is consistent with the way the subject is really interpreted - as VP, which means we are not talking about planes that land[6]. 'Bylo' is singular neuter, even though there is no singular neuter item to agree with, just as in Russian quantified subject cases we analyzed before.

Fujita (2017) argues that there are no "flavors" of ν and suggests that the transitivity alternation reduces to the choice of label. Unlike in traditional view of transitive structure (18a), ν (which is always strong) and V first merge to form {ν,V} as in (18b). Then, depending on what is α, the label, the structure becomes either transitive or ergative/unaccusative. ν would fail to license EA when α is V. It might also become "invisible" and thus stop functioning as a phase-head, as in the similar approach by Epstein et. al (2016).

[6] In Russian, the subject is unambiguous even though accusative case marked 'samolëty' looks exactly the same as its nominative form. This also strengthens the structure-dependent agreement view.

(18)



2.4. Summary

In general, the evidence strongly suggests that NP subjects successfully agree with verbs and QP subjects fail to agree. In case there is non-agreement, both Case and φ-feature agreement fail. Given the data above and assumptions we made in the introduction section, we can make the following, stronger generalization:

(19) X can be involved in agreement operations iff X is a label.

It is surprising that QP subjects that fail any kind of agreement still occur in Spec/TP. This implies that movement of subjects to Spec/TP is not motivated by feature checking itself. However if Internal Merge is indeed free and QPs are not triggered by T, then there is little surprising in this fact. QPs move blindly and either agree or fail to agree. Something like EPP might be at work, however it is still unclear why EPP should exist. EPP itself might be reduced to labeling requirements, but it is unlikely that anything like feature sharing takes place in Spec/TP, especially if there are *no* features to share as is the case with quantified subjects.

Throughout the section I assumed that whenever N in {Q,N} fails to agree, it is because Q by virtue of being the label blocks agreement. However (19) *per se* allows the possibility that *none* is the label. I deliberately leave this possibility open. To block it would mean to claim that every structure *must* be labeled. While on purely conceptual grounds this seems reasonable, stronger evidence is required. If label is needed at the interface in the spirit of SMT, then evidence could be obtained by investigating whether something like label exists in other cognitive faculties, with which language faculty interacts.

3. The nature of labels

3.1. The role of labeling in narrow syntax

Merge is an operation that takes two objects and forms a larger unit out of them. It is taken for granted that each application of Merge yields larger and larger *nested* structures. However this view relies on a hidden premise that Merge, as defined *over atoms*, automatically yields a mergeable atom. In fact, Merge need not have been defined so. Simplest Merge would take X and Y, each an atom, and form {X,Y} but is {X,Y} really *atomic*? Merge *could* be a recursive set union operation that does form

larger units but instead of yielding nested structures like {Z, {X, Y}} it would yield {X,Y, Z}[7]. If we assume that Merge is defined over atoms but the product {X,Y} is not an atom then Merge would fail to apply recursively. What would make {X,Y} atomic is a function that maps the set onto one of the members, that is, X or Y[8]. This operation is Label and it yields either X or Y, which is atomic and thus can be an input to recursive applications of Merge. The label of the set is one of the atoms themselves, allowing for a bare phrase structure. This is essentially the reasoning in Hornstein (2009)[9]. If this is on the right track, then labeling not only enables interpretation of syntactic objects but also makes recursive structure building possible. It is then reasonable to assume that labeling should apply immediately - otherwise there would be no structure.

3.2. Labeling as symmetry breaking in other cognitive faculties

What is the nature of labeling? Label clearly influences the way we interpret syntactic objects, so it is reasonable to believe that labeling could be nothing but a "solution" that makes objects generated by Merge interpretable at all. If labeling is required to make syntactic structure interpretable, then we might also ask whether it is required to interpret "non-linguistic" objects - information on which other cognitive faculties operate. By "being interpretable" we usually mean "being usable by other cognitive faculties" when we speak about the objects of the narrow syntax. However, if something akin to labeling is found in other faculties as well, we might conjecture that labeling makes *anything* usable *anywhere* in the mind, thus literally enabling complex thought and cross-modal interaction. Without labeling, then, it would be hard e.g. to draw or talk about what we see, plan actions, enjoy music instead of just hearing notes as random sounds, use symbols to indicate direction, and maybe even orient ourselves in space. Let us see how plausible this conjecture might be. We can reasonably start asking whether spatial cognition or temporal cognition also utilize something akin to labeling. One example which I believe is relevant is temporal (and causal) relationships.

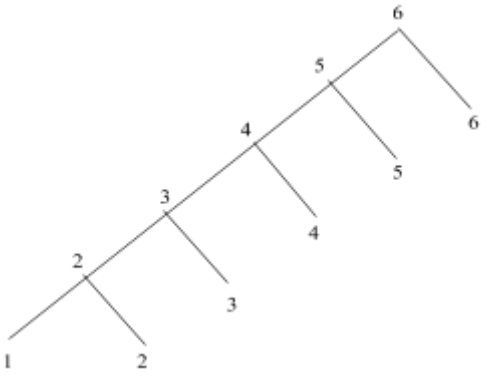
3.2.1. Temporal cognition

Humans are capable of understanding that one event precedes or follows another in time. Of course, we also have no problem with understanding that an event follows a *set* of other events in time. To make this more complex ability possible, it is not enough to just establish a local relationship between two last events of the complex event chain. Otherwise we would not know whether the last event of e.g. six consequent events precedes or follows the first one. How can this be achieved in the simplest way? Suppose that relationship is always established locally, between two temporally adjacent events. But once time values (whatever their nature is) are compared, the set of these two events is labeled with the time value of the last event. Then, when the next

[7] Strictly speaking there is a difference between set union and e.g. ternary Merge, because the latter does not require one of the arguments to be a set.
[8] Or it could be an exocentric structure, with neither X nor Y functioning as the label.
[9] Hornstein analyzes Merge as a complex operation, which consists of Concatenate and Label. But if Simplest Merge really yields non-atomic objects then there seems to be no significant difference between it and Concatenate as defined in his work.

event happens the relationship will again be established locally but this time between the last event and the *label* of the set of preceding events. The structure for some particular chain of arbitrary 6 events can be represented as (20)[10], where numbers are just simplified numerical time values (milliseconds, perhaps):

(20)



We say that 1 is *contained* in 2, and 2 is contained in 3 etc. Due to this, the relationship between 6 and 1 can be established transitively by establishing it locally between 6 and 5. Without labels this structure would be meaningless - the only information available would be that 6 and the set of other sets of events are members of one larger set but no particular relationship (apart from sisterhood) between these members is established. Moreover since members are not ordered a big amount of different branchings would be possible. The interpretation of causal relationships would be completely ruined (or rather, be *very* different from how humans do it). It is also possible that each set is labeled as "before" - then for the last event the set of other events will always be "before"[11]. Interestingly, if this is indeed the structure of subjective time then some minimal search algorithm would almost suffice for labeling. But even in this case the relationship between 1 and 2 would not be established which is not the result we want.

Note that the structure above is binary. It does not have to be so, but let us imagine what happens when it is non-binary. Binarity allows us to make a very simple comparison like "3 is greater than 2". How would we compare a set of three values, say, {V1, V2, V3}? In the present context this would also mean that we do not establish the relationship immediately and "wait" for more events. But how long should we wait? Should it be the third event? Why not the fourth or the fifth? It is unclear how waiting could contribute to interpretation, and such algorithm would face the problem of establishing the global relationship between multiple events.

Interesting evidence comes from Eagleman & Sejnowski (2000) (ES). They investigate the nature of the so-called flash lag illusion, wherein a flash occurs *within* a moving circle, but the circle is perceived *further* in distance than the flash (Fig.1). ES

[10] The similarity of the structure to that of syntax should be taken to be accidental, because I do not claim that something like Merge is responsible for constructing it.

[11] However this would make the structure exocentric and imply that the concept of "before" is not dependent even on such simple operation as labeling.

argue that *extrapolation* and *latency difference* interpretations of flash lag illusion are inconsistent with new experimental data. This means we misperceive the position of the circle not because we are capable of predicting its future positions and not because *every* single event is processed with delay. ES explain the illusion by *postdiction* - there is an approximately 80ms window of events in *real* time, which are subjectively processed by humans as "now" *at once.* This means that the moment of flash and the position of the circle 80ms after the flash are all processed as "now" which causes the illusion.

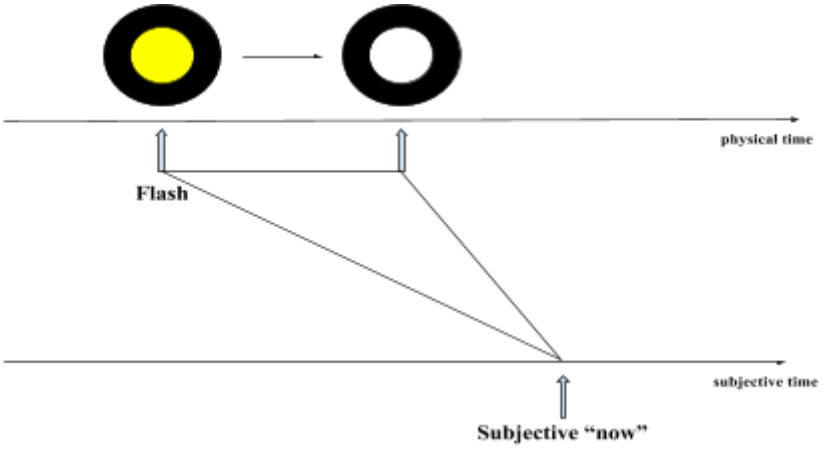


Fig.1. Flash lag illusion.

Due to the way our body is designed, even physically simultaneous signals may take different time to travel through the nervous system (nose is closer to the brain than toe and auditory cortex responds to stimuli faster than visual cortex). Brain "waits" and accumulates signals from different sensory modalities to perceive them as "now" (Stetson et. al, 2006). Suppose that this accumulation is nothing but structuring (sub-)events as in (20) above for ~80ms. Then, due to the fact that every real event *is* represented in the structure by being stacked, it is simply impossible to perceive the circle and the flash in the same position when we interpret (20) as "now", because of the intervention of other events in the structure. A hypothetical brain that simply delays processing of every single event would not experience this effect due to the absence of accumulated structure.

3.2.2. Spatial cognition

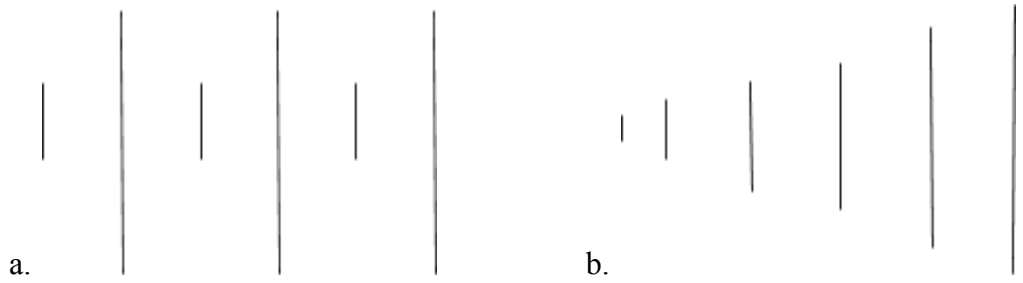
Next, let us take a look at an example from visual cognition. Because of properties of light we perceive the *angular size* of objects, that is, see in perspective. It is also known that such phenomena as binocular disparity or motion parallax enable depth perception (Gibson 1959, Marr 1982). However the ability to perceive depth from different clues allows us to infer the *real* (absolute) size of objects that are at different distances from our eye and to infer the distance when the absolute size is previously known and, in some cases, even when it is unknown (Sousa et. al (2011)). This is how we understand that even if there are two similar houses along the street and one of them seems bigger,

it is because it is closer. The clues influence perception so strongly that they can force us to believe that an object is bigger than it really is, as in Ponzo illusion. In general, it seems that not only depth from stereopsis or parallax can help us infer the size, but the size itself can be used to infer the distance/depth when other clues are lacking. In fact, this becomes clear even with a simple thought experiment.

Imagine that in front of you is an *ideal* painting of a street with street lamps along it. The painting is so big that you do not see its borders and so realistic that you cannot know whether it is a painting until you touch it or turn your head. In this case, such clues as horizontal disparity or motion are absent since neither we nor the painting move, and every dot on the plane is at the same distance from our eyes unlike in a 3D scene. Yet we are able to perceive depth as if it was a real street. Note one obvious, but important thing - real street lamps are constant in size, they are *not* smaller or bigger than each other. In the painting, however, their sizes are *different* - one lamp is actually drawn smaller or bigger than another. What does it tell us? I believe that it confirms that in the absence of any 3D clues, just size can be enough to *fully* perceive depth, perhaps even when the objects are unfamiliar. Interestingly, when asked to draw two cubes in perspective, a bad drawer would simply draw a bigger and a smaller cube next to each other with their edges on parallel lines, while more correct representation requires them to be drawn along two lines that converge at a vanishing point. A lot of ancient paintings exploited this less geometrically correct and naive understanding of perspective. Yet it seems that this is because we perceive the world this way.

The size clue is so simple that we can even ask what kind of operations are involved in establishing the relationship "closer/further than" and, more generally, "to be in perspective". Imagine that we have 6 lamps. In absence of any other clues, we say that 6 lamps are in perspective only if in each pair of the lamps, one is longer than another. However, this much is not enough since we do not say that lamps (as lines) are in perspective in the picture (21a) below, while they satisfy the condition just given.

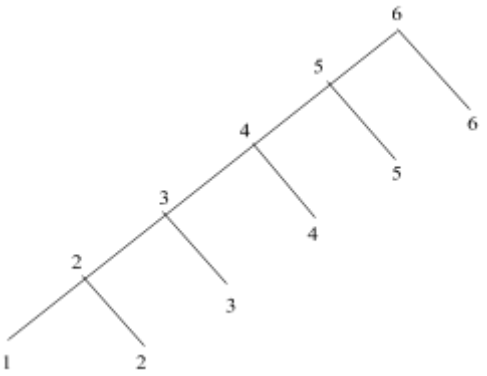
(21)



a. b.

Note that establishing only *local* relationships may give us (21b) too, but it will be only one of many possible combinations. However, we recognize *only* (21b) as "being in perspective". Thus having (21b) just as *one* of the many possibilities is simply not consistent with how humans perceive depth (but other creatures *might*). To make *only* (21b) possible, we need that every line be longer than the longest line in the *set* of lines that we already compared. The structure of relationships for this particular case can be simply represented as (22) which is identical to (20):

(22)



Numbers here are numerical values of lengths. Just like in (20) the *global* relationship between 1 and 6 can be established transitively through labels which make containment possible. Without labels, the structure would again be meaningless and possibly fit interpretations like (21a) which have nothing to do with perspective. Moreover, this structure must also be binary - this is the simplest way to avoid (21a)-like interpretations and any non-binary variant would require more complex algorithms. There is of course very little surprising in the fact that we are using 2D, impoverished information to understand a 3D scene (consider recovery of the 3D structure from motion by rigidity principle (Ullman 1979)). *If* labeling allows for depth perception in this particular case then we can expect organisms that lack labeling to have difficulties with orientation in space when not many other clues are available.

Strictly speaking, labeling in examples above serves to make a meaningful (or perceptual) *distinction* between two objects in question. The set of two lines as in the example above is just what it is - two lines of different lengths together do not mean anything, but once label creates a distinction between them, the set can serve as usable (and potentially useful) information. It does not *have* to be used to infer depth - but in human case it is exactly how it seems to be used. In the case of language, we could say that {V, N} is just a set, but only once the set is labeled as V, the noun *becomes* the verb's object, *gaining* a semantic function that was absent before. Then it makes no sense to talk about selection, which is rather an epiphenomenon of labeling. We will return to this later. Another domain where objects seem to gain function only under labeling is musical cognition.

3.2.3. Musical cognition

Without absolute pitch and education in musical theory, one would not be able to correctly name the note (such as Cb or G), the interval (such as minor third) or the scale (e.g. Eb major) being played. But the difference is nevertheless perceived by everyone and certain combinations of notes affect a listener in a certain way. Musical theorists have names for different musical constructs and this obviously means that these constructs are perceptually distinctive. One important notion is that of *function*. A note,

for example, is the basic fundamental block which is used to create music. However a note alone is "useless" until it is put in some musical context. Although notes can constitute scales, a single note played does not belong to any scale in a natural sense. For example, G can be part of both C major or G minor, but we perceive it as belonging to one of these scales only after we hear other notes in the context. This is because G or any other note can serve different function(s) in different scales (e.g. tonic, subdominant, dominant, leading etc.). The function is calculated against the root, the "main" note of the scale - the *ratio* of frequencies of the note and the root, in particular. For example, in C major, G is 7 half-steps away from C, the root. The *interval* they form is called perfect fifth and it corresponds to a certain frequency ratio (3:2), to which human listeners are sensitive, along with many other intervals such as octave 2:1, major third 5:4, unison 1:1 etc. In G minor, however, G itself is the root and it does not form a perfect fifth with C, but does so with D. This way, the root determines the function of every note in the scale. The perception of the scale can be thought of as a process of constructing a mental structure in which notes are related in a certain way. But this is possible only because we are capable of calculating frequency ratios, and this is what allows us to perceive the root of the scale.

What happens if such a structure cannot be labeled? Can we expect it to be uninterpretable/unusable or at least not usable in the same way as structured scales are? It turns out that the answer is yes. One interesting example is the so-called *whole tone* scale. The peculiar property of it is that all notes are one whole step away from each other. Due to this, some intervals (calculated frequency ratios) such as major third can be identified *multiple* times but with different notes. This leads to the impression that virtually every note of the scale can function as the root. Since the root is exactly what helps us determine the scale, the listener never knows which scale he is in when he hears a piece which uses the whole tone scale[12]. It is sometimes said to be "ambiguous" or "lack direction and center" as opposed to scales where each interval can be identified a *unique* number of times (e.g. major or minor). Having each interval a unique number of times is referred to as a "deep scale property", which the whole tone scale lacks (Johnson, 2003). Music that uses this kind of scales is called *atonal* and it affects listeners in a strikingly different way. Interestingly such scales are called *symmetric* scales, as opposed to *asymmetric* scales such as major scale or minor scale. Patel (2010) reports:

> "Trehub et al. (1999) constructed unfamiliar asymmetric and symmetric scales based on 7 intervals per octave, and tested the ability of infants and adults to detect subtle pitch changes in repetitions of these scales. Infants were better at detecting such changes in the asymmetric scales, even though both scales were unfamiliar, suggesting that the asymmetry conferred some processing advantage [...] Trehub et al.'s results may be relevant to the origin of musical scale structure in diverse cultures, because in contexts in which there is no "familiar" scale, the cognitive system may be biased toward choosing an asymmetric scale structure over a symmetric one because of a predisposition for music with a clear tonal center."

[12] Such scales as the whole tone scale are also called "modes of limited transposition" (Messiaen 1956).

While symmetric scales certainly do affect a listener in *some* way and are used in experimental pieces, notes within it lack any specific function due to the absence of perceived root.

To summarize, we can say that tonal scales are structures labeled[13] by frequency ratio calculation, as a *specific* algorithm of the musical faculty serving the *general* requirement on usability/interpretation. To avoid a possible confusion that some operation Label is universal across the cognitive faculties, we might refer to the operations as *symmetry breaking*. Such operations are likely to be domain specific and exploit (perceptual) distinctions available only to one domain but they seem to serve the same general requirement on usability/interpretability by doing essentially the same thing - establishing asymmetry between two objects. Were the operation domain-general it would have to "know" how to label all kinds of objects on which cognitive faculties operate - e.g. how to compare numerical values, how to compare frequencies and how to choose verb over noun and vice versa. It does not seem to me very plausible. Whenever asymmetry cannot be established, the objects in question do not seem to be semantically interpretable in case of language, or useful for spatial cognition etc. It is then important to ask - why does our mind have to be this way? What about it makes it function the way it does only if objects on which each faculty operates are asymmetric? Perhaps there is a species-specific universal interpretive faculty, that requires everything that reaches it to be asymmetric. Since language does not come from the outer world in the shape of e.g. lines but comes from the "inside", it must have taken the form that is possible to label to satisfy general interpretation requirements. Interestingly, Merge provides exactly such objects, so we could say that Merge itself is the perfect "solution".

3.3. Evolutionary context

Hornstein (2009) believes that labeling is *the* evolutionary innovation. He writes:

> "Logically, hierarchy and recursion are independent of labeling. And not only logically: for example, it has been observed that there is nesting in language without endocentric labeling, as in, for example, syllables which have [onset [nucleus coda]] structure. There is no endocentric labeling in syllables and, interestingly, we do not find repeated nesting in such configurations, i.e. syllables within syllables."

This is a very important point, because it means that simply being able to create nested structures is not sufficient to give rise to natural languages humans have. Take the observed similarity between syntactic structures and the way chimpanzees use tools. The behaviour can be represented as e.g. [hammer [nut anvil]] (Fujita 2014). Even if this is indeed the structure of the chimpanzee's behaviour and even if it is constructed by something like Merge, it lacks labels[14]. If labels are required to make natural languages

[13] Also see Fujita (2017) for labeling in chord inversion and the way Merge strategies might influence labeling.

[14] See Berwick & Chomsky (2017) for the claim that stone tool making does not involve labeling and therefore it is not related to merge/syntax.

and (linguistic) thought possible then it is not very surprising that chimpanzees do not possess language.

However, it is not very clear how labeling can benefit an organism if it only enables natural languages. I believe that labeling, as a more abstract instance of symmetry breaking, not just leads to emergence of language, but significantly improves other modes of cognition, such as visual or temporal cognition as is shown in the previous sections. Clearly such a "boost" in cognitive abilities across all domains would significantly benefit species. We can conjecture that labeling "solutions" emerge automatically once there is a usability/interpretability requirement imposed by some universal interpretive mechanism. Then it would be the emergence of this very mechanism that "boosts" multiple modalities and gives enormous advantage to species. Language could have arisen in our species as a by-product of satisfying general requirement on interpretability/usability which labeling of syntactic structures makes possible.

4. Conclusion

In this paper I argued against approaching the problem of labeling with minimal search algorithms and suggested that in fact lexical items are fully responsible for label determination. I showed how label affects the workings of narrow syntax by focusing on unusual agreement phenomena in Russian. Labels seem to block φ-feature agreement of the object that is not the label as in {Q,N} or {V,N} whenever Q or V is the label, respectively. Labels can also explain other phenomena such as transitivity alternation or phase-cancellation.

I investigated the status of labeling within temporal, visual and musical cognition and argued that the purpose of labeling (in a broad sense) is establishment of asymmetry, which allows us to make perceptual/meaningful distinctions about the world and to make objects on which a cognitive faculty operates usable in other domains. Labeling seems to be not only "conceptually necessary" but also empirically necessary for objects generated by narrow syntax, which have to be interpreted at the interfaces that relate language to the rest of the cognitive faculties, just like objects of other domains. Though I could identify only few cases where labeling seems to matter, there could be many more waiting to be discovered. The question of what makes our mind work exactly this way is left without any satisfactory answer and requires more investigation.

References

Bailyn, John F. (2012). *The Syntax of Russian*. Cambridge University Press.

Berwick Robert C., Chomsky N. (2017). Why only us: Recent questions and answers. *J.Neurolinguistics*, 43(B): 166-177.

Boskovic, Zeljko. (2003). A Minimalist Account of Genitive of Quantification, paper presented at Formal Descriptions of Slavic Languages 5, Leipzig.

Bošković, Ž. (2006). Case and Agreement with Genitive of Quantification in Russian.

Chomsky, N. (1994a). Bare Phrase Structure.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (1998). *Minimalist inquiries: The framework* (No. 15). MIT Working Papers in Linguistics, MIT, Department of Linguistics.

Chomsky N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge, UK; New York: Cambridge University Press.

Chomsky, N. (2008). On phases, In *Foundational Issues in Linguistic Theory* ed. by R. Freidin, . Otero and M.-L. Zubizarreta, 136-166. Cambridge MA: MIT Press.

Chomsky, N. (2013a). Problems of projection. *Lingua* 130: 33-49.

Chomsky, N. (2015a). Problems of projection: Extensions, In *Structures, strategies and beyond: Studies in honor of Adriana Belletti*, ed. by E. D. Domenico, C. Hamann and S. Matteini, 3-16. Amsterdam: John Benjamins.

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, 287(5460), 2036-2038.

Epstein, S. D., Kitahara, H., & Seely, D. (2016). Phase cancellation by external pair-merge of heads. *The Linguistic Review,* 33(1): 87-102.

Etxeberria U. , Etxepare R. (2012). Number agreement in Basque: counting vs. measuring, In *Noun Phrases and Nominalizations in Basque* ed. by Etxeberria, U., R. Etxepare, Uribe-Etxebarria, M. Amsterdam: John Benjamins.

Fox, D. (2000). *Economy and Semantic Interpretation*. Cambridge, MA: MIT Press.

Franks, S., Hornstein, N. (1992). Secondary predication in Russian and proper government of PRO, In *Control and Grammar*, ed. by R. K. Larson et al., 150, Dordrecht: Kluwer.

Franks, Steven. (1994). Parametric Properties of Numeral Phrases in Slavic. *Natural Language and Linguistic Theory* 12: 597-674.

Franks, Steven. (1995). *Parameters of Slavic Morphosyntax*. Oxford University Press.

Fujita, K. (2014). Recursive Merge and human language evolution. In *Recursion: Complexity in cognition* (pp. 243-264). Springer International Publishing.

Fujita, K. (2017). On the parallel evolution of syntax and lexicon: A Merge-only view. *J.Neurolinguistics*, 43(B): 178-192.

Gibson, E. J., J. J. Gibson, O. W. Smith, and H. Flock. (1959). Motion parallax as a determinant of perceived depth. *J. E. Psychol*. 8, 40-51.

Hornstein, N. (2009). *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.

Ionin, T. (2001). Scope in Russian: Quantifier Movement and Discourse Function. Ms., MIT.

Johnson, Timothy A. (2003). Foundations of Diatonic Theory: A Mathematically Based Approach to Music Fundamentals. Mathematics Across the Curriculum. Emeryville, CA: Key College Publishing.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH San Francisco: Freeman and Company, 1(2).

Messiaen, O. (1956). The Technique of My Musical Language, trans. *John Satterfield (Paris: Alphonse Leduc, 1956)*, 46.

Neidle, C. (1988). *The Role of Case in Russian Syntax*. Dordrecht: Reidel.

Paperno, D. (2012). Quantification in Standard Russian, In *Handbook of Quantification in Natural Language* ed. by Edward L. Keenan, Denis Paperno. Springer, 2012.

Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.

Pereltsvaig A. (2006). Small nominals. *Natural Language & Linguistic Theory* 24: 433-500.

Pesetsky, D. (1982). Paths and Categories. PhD Dissertation, MIT.

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, 18(3), 359-368.

Sousa, R., Brenner, E., & Smeets, J. B. J. (2011). Judging an unfamiliar object's distance from its retinal image size. Journal of Vision, 11(9), 10, 1-6.

Sousa, R., Smeets, J. B. J., & Brenner, E. (2012). Does size matter? Perception, 41(12), 1532-1534.

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, 51(5), 651-659.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, 25: 965-975.

Ullman, S. (1979). *The interpretation of visual motion*. MIT Press.