Homophony and Russian copular constructions

Erin Zaroukian (Johns Hopkins University)

Certain languages, notably Russian, appear to ban consecutive homophonous *wh*-phrases. It turns out, however, that this ban extends beyond *wh*-phrases. Here I provide a Distinctness-based analysis under which banned sentences are those which result in an unlinearizable node pair. This analysis has ramifications for the structure of Russian copular constructions, in particular suggesting that an overt copula introduces a phase head, but a null copula does not.

1. The data

Bošković (2002) notes that while Russian is a multiple *wh*-fronting language, multiple *wh*-fronting does not occur if it would result in consecutive homophonous *wh*-phrases, as shown in (1).

(1) a. *Čto čto obuslovilo? (Bošković 2002:364)
what what conditioned
b. Čto obuslovilo čto?
what conditioned what
'What conditioned what?'

Bošković's response to this anomaly is a "...PF constraint against consecutive homophonous *wh*-phrases..." (Bošković 2002:365). In this paper I show that the data is more complicated than (1) suggests, and I argue that this PF constraint alone will not suffice to capture the overall pattern.

Consider first the null copula, which is frequently utilized in Russian, as demonstrated in (2).

(2) Kto (?est') Ivan?
who (is) Ivan
'Who is Ivan?'

If a copular sentence consists of two homophonous *wh*-phrases, failure to front one of the *wh*-phrases would still result in consecutive homophonous *wh*-phrases, as can be seen in (3a). Here, pronunciation of the copula is obligatory, shown in (3b). This is consistent with Bošković's PF constraint, since the overt copula causes the homophonous *wh*-phrases to no longer be consecutive.

(3) a. *Kto kto?
who who
b. Kto est' kto?
who is who
'Who is who?'

Now consider the following cases where the consecutive homophonous elements are not *wh*-phrases. In (4) we see obligatory copula pronunciation with homophonous nouns.[1]

(4) * *noun – noun*
a. *Vrač vrač.
doctor doctor
b. Vrač est' vrač?
doctor is doctor
'The/a doctor is the/a doctor?'

In (5), even though there are no consecutive homophonous nouns, copula pronunciation is still obligatory, making it appear that consecutive homophonous adjective-noun pairs are likewise banned.

(5) * *adj-noun – adj-noun*
a. *Bogatyj vrač bogatyj vrač.
rich doctor rich doctor
b. Bogatyj vrač est' bogatyj vrač?
rich doctor is rich doctor
'The/a rich doctor is the/a rich doctor?'

In (6) we see that, unlike the previous instances, consecutive homophonous adjectives are allowed.

(6) *adj – adj*
a. Bogatyj bogatyj.
rich rich
'The rich person is rich.'
b. Bogatyj bogatyj vrač?
rich rich doctor
'The rich person is a rich doctor?'

[1] The insertion of emphatic *že* may be another repair strategy that will not be explored here.

Similarly, consecutive homophonous verbs in (7) and adverb-verb pairs in (8) are allowed.

(7) *verb – verb*
Emu nravitsja nravit'sja.
he likes to-be-liked
'He likes to be liked.'

(8) *adverb-verb – adverb-verb*
Emu umyšlenno nravitsja umyšlenno nravit'sja.
he deliberately likes deliberately to-be-liked
'He deliberately likes to be deliberately liked.'

A summary of this data is provided in (9).

(9)

Acceptable	Banned
adj – adj (6)	*wh* – *wh* (1),(3)
verb – verb (7)	noun– noun (4)
verb phrase – verb phrase (8)	noun phrase – noun phrase (5)

The data presented above suggests that Bošković's PF constraint should extend beyond *wh*-phrases. As may be apparent from the table in (9), a generalization over banned instances can be gleaned, and we might reformulate the PF constraint as one not against just consecutive homophonous *wh*-phrases, but as one against consecutive homophonous noun phrases. It is not clear, however, that restating the data like this is the ideal way to account for it.

2. Distinctness

Repetition avoidance is quite widespread and can been seen in domains from phonology to syntax to discourse structure (Walter 2007). For example, in phonology the Obligatory Contour Principle (OCP) is invoked to account for patterns such as the lack of successive occurrences of [+labial] segments in certain languages (Fukazawa 1999). Similarly, the double *-ing* filter in English is used to account for the ungrammaticality of sentences like **It's continuing raining* (Ross 1972).

A desire for parsimony might lead us to seek a unified explanation behind all linguistic repetition avoidances, but Walter (2007) suggests that there are at least three different reasons to avoid repetition, drawing from physiology, perception, and *Distinctness*. It is the third of these that is relevant to syntax, and it will be

explored below as an explanation for the pattern of repetition avoidance seen in (1)-(8).

2.1 Richards (2010) on Distinctness

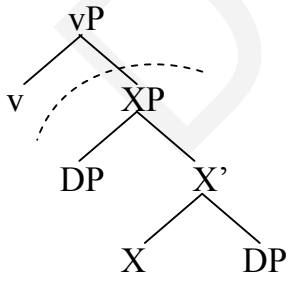
Richards (2010) seeks to provide a unified account for syntactic repetition avoidance. The pattern he remarks on is that multiple objects of the same type cannot occur too close together, and to account for this he proposes a well-formedness condition on linearization such that multiple syntactic nodes of the same kind cannot form a linearization statement, which he terms *Distinctness*.

(10) Distinctness: If a linearization statement <α,α> is generated the derivation crashes (Richards 2010:5)

Richards assumes the LCA (Kayne 1994), where the pair <α,α> is determined by asymmetric c-command. He also assumes that linearization proceeds in phases, with the strong phases including CP, transitive vP[2], PP, and KP. Additionally, he assumes that only (features that percolate up from) functional heads give rise to distinctness violations.

To see Distinctness in action, we can first assume that linearization only has access to node labels. In (11), the sister of v is spelled out (indicated by the dashed line), generating the linearization statement <DP,DP>. This statement of the form <α,α> will cause the derivation to crash.

(11) <DP, DP>✖



Some banned English structures with this general form are given in (12) below. These show Distinctness violations through multiple sluicing (Richards 2010:8, cf. grammatical *I know everyone danced with someone, but I don't know who with whom*), multiple exceptives (Richards 2010:8, cf. grammatical *Every man*

[2] None of the Russian data discussed here will depend on intransitive vP not being a phase. See Legate (2003) for details.

danced with every woman, except John with Mary), and quotative inversion (Richards 2010:13, cf. grammatical *"It's raining", said John (to Mary)*).[3]

(12) a. *I know everyone insulted someone, but I don't know who who(m).
b. *Every man admired every woman, except John Mary.
c. *"It's raining", told John Mary.

In Russian, the story becomes slightly more complicated. Given the examples in (13) and (14), Richards determines that linearization in Russian, unlike English, has access to gender and case features. It then follows that when two DPs are linearized together, the derivation crashes only if they cannot be distinguished though gender and case features, as in (13b), shown in (14b).

(13) a. Kakomu žurnalistu kakogo diplomata nužno zavtra privetstvovat'?
which$_{dat}$ journalist$_{dat}$ which$_{acc}$ diplomat$_{acc}$ must tomorrow greet?
'Which journalist needs to greet which diplomat tomorrow?'
b. ??Kakomu žurnalistu kakomu diplomatu nužno zavtra zvonit'?
which$_{dat}$ journalist$_{dat}$ which$_{dat}$ diplomat$_{dat}$ must tomorrow call?
'Which journalist needs to call which diplomat tomorrow?'
c. Kakomu žurnalistu kakoj ženščine nužno zavtra zvonit'?
which$_{dat}$ journalist$_{dat}$ which$_{dat}$ woman$_{dat}$ must tomorrow call?
'Which journalist needs to call which woman tomorrow?'

(14) a. [$_{FP}$[$_{DP[dat,masc]}$ Kakomu žurnalistu] [$_{FP}$[$_{DP[acc,masc]}$ kakogo diplomata]
[$_{IP}$ nužno zavtra privetstvovat']]] <DP$_{[dat, masc]}$, DP$_{[acc,masc]}$>✓
b. [$_{FP}$[$_{DP[dat,masc]}$ Kakomu žurnalistu] [$_{FP}$[$_{DP[dat,masc]}$ kakomu diplomatu]
[$_{IP}$ nužno zavtra zvonit']]] <DP$_{[dat, masc]}$, DP$_{[dat,masc]}$>✗
c. [$_{FP}$[$_{DP[dat,masc]}$ Kakomu žurnalistu] [$_{FP}$[$_{DP[dat,fem]}$ kakoj ženščine]
[$_{IP}$ nužno zavtra zvonit']]] <DP$_{[dat, masc]}$, DP$_{[dat,fem]}$>✓

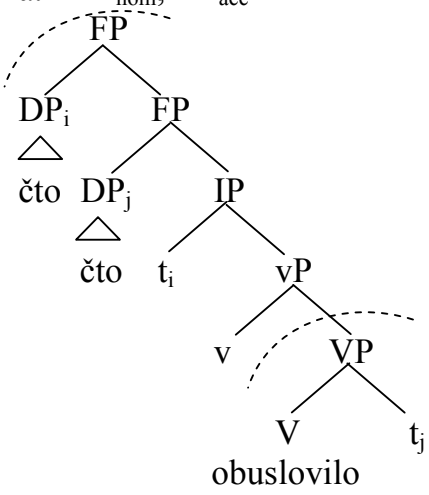
2.2 Applying Distinctness

In what follows, a Distinctness analysis is applied to the data introduced above. The data in (1) is shown in (15). In (15a), the two DPs are spelled out in the same phase, which we hope would lead to a Distinctness violation and explain the sentence's ungrammaticality. These DPs, however, differ in case (nominative and accusative) and should therefore be distinguishable at linearization. Though as Richards notes, case syncretism plays a role at linearization. We will

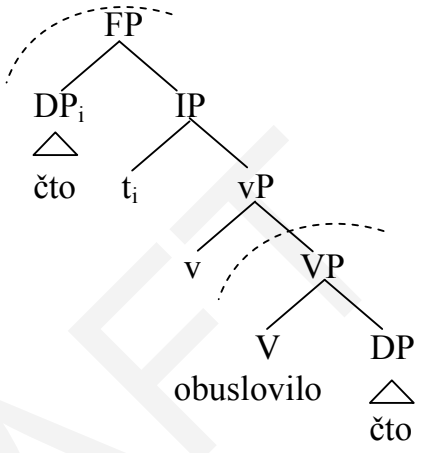
[3] Though these examples all show items being string adjacent, Richards provides evidence that string adjacency is neither necessary nor sufficient. For example, the insertion of an adverb (e.g. *quickly*) between the two DPs in (12c) does not improve acceptability.

assume, following Richards and Distributive Morphology (Bonet 1991), that case syncretism involves impoverished features (symbolized by the strike-through in (15)) which do not allow case-syncretic elements to be distinguished at linearization. Now in (15a) two indistinguishable DP are linearized in the same phase, resulting in a Distinctness violation. In (15b), however, one of the DPs remains in situ and is spelled out in a lower phase, causing no problems for linearization.

(15) a. <$DP_{\sout{nom}}$, $DP_{\sout{acc}}$>✖

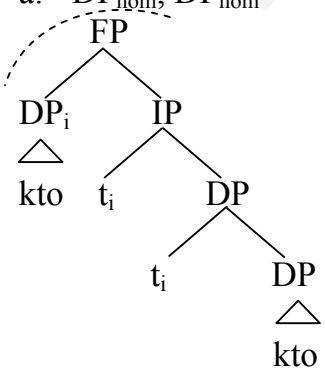


b. ✓

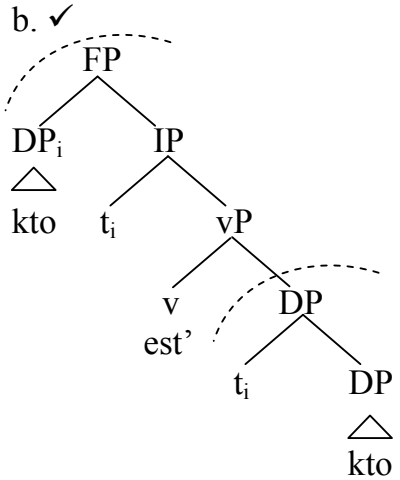


For the data in (3), I will assume that the overt copula is a light verb and that constructions with null copulas lack a vP layer (these assumptions will be discussed below). It then follows that (16a), where two nominative DPs are spelled out in the same phase, should crash. The insertion of a copula introduces an additional spell-out phase such that the two nominative DPs do not directly form a linearization statement.

(16) a. <DP_{nom}, DP_{nom}>✖

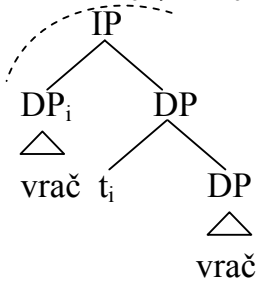


b. ✓

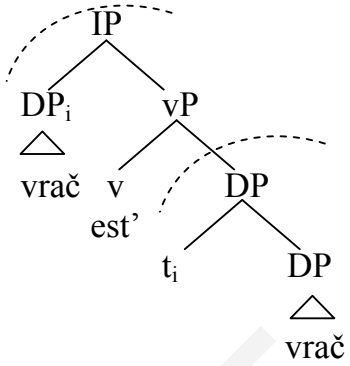


The same pattern from (4) can be seen in (17), though now with non-*wh*-phrases.

(17) a. <DP_{nom}, DP_{nom}>✗

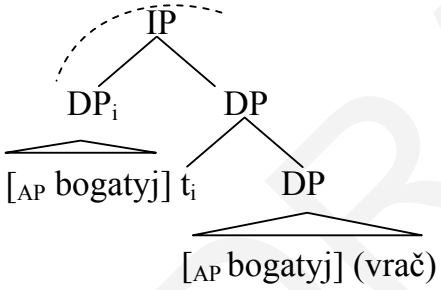


b. ✓



At first blush, this Distinctness analysis fares less well with adjectives. Though the sentences in (6) are grammatical, (18) suggests that they should result in an unlinearizable pair.

(18) <DP_{nom}, DP_{nom}>✗!

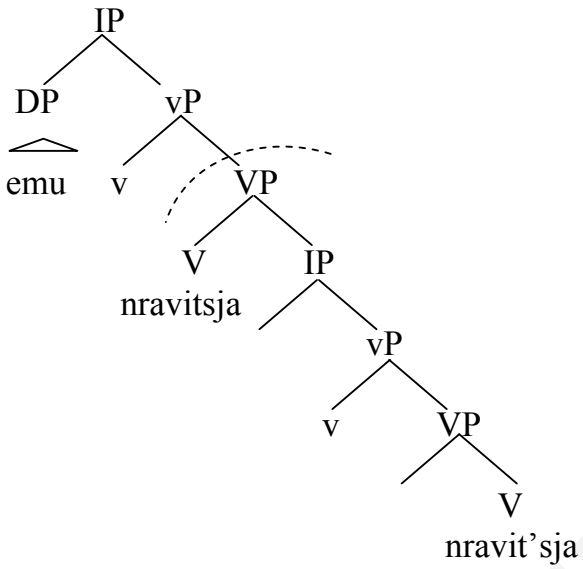


It appears, then, that the two homophonous adjectives must differ with respect to some relevant feature which is visible at linearization, making them distinct.

A similar issue arises with verbs. The sentence from (7) is drawn in (19), which is grammatical despite the linearization statement <v,v>[4]. This suggests that linearization in Russian is sensitive to person/number features which cause the two vs to be distinct.

[4] Here I treat both verbs as intransitive. If *nravitsja* 'likes' were treated as transitive, the vs would be spelled out in separate phases, creating no trouble for linearization.

(19) <v_{3ps},v_{inf}>



So, a Distinctness analysis can be used to predict the pattern of judgments in (1)-(8). The picture becomes more complicated, however, when data like (20) and (21) (cf. (4) and (5)) are considered. Here the DPs remain homophonous and share the same case and gender features, but a semantic contrast appears to make them linearizable.

(20) Vrač vrač.
doctor doctor
'Mr. Vrač is a doctor.'

(21) Bogatyj vrač bogatyj vrač.
rich doctor rich doctor
'The (financially) rich doctor is rich (in that he has wonderful friends).'

The most obvious solution would be for Russian linearization to be sensitive to yet more features, which would be tied to these semantic contrasts. Further research would be needed to confirm this hypothesis.

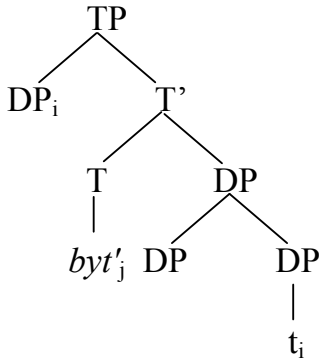
The richness of features called for here might seem suspicious, but similar richness is independently called for under this analysis to account for null copula sentences like (2). In (20), (21), and (2) the sharper semantic contrast between the DPs, if captured with syntactic features visible at linearization, accounts for their grammaticality.[5]

[5] Equative and specificational sentences also seem to rely on a high sensitivity to semantic contrasts, where (ia) does not require an overt copula, though the noun phrases match in case and gender. These constructions, however, will be largely ignored in the interest of space. This issue of feature visibility will be returned to in the conclusion.

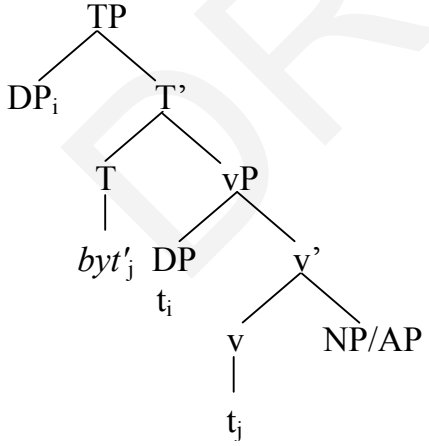
2.3 Consequences for copular constructions

Russian copular constructions are quite complex and typically receive more semantic than syntactic attention. Notable among syntactic analyses, Pereltsvaig (2001) suggests the following structures for Russian non-existential copular constructions.[6] The structure in (22a) is used in constructions where the second element appears in nominative case (as it has in the examples discussed above), and the structure in (22b) is used in constructions where the second element appears in instrumental case.

(22) a. Bare Copular Sentences (nominative)



b. Rich Copular Sentences (instrumental)

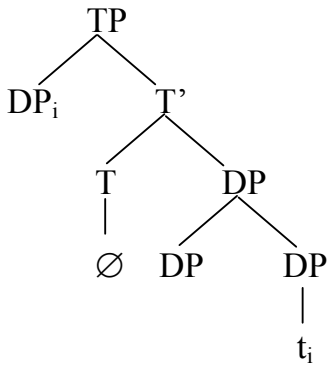


(i) a. Vrač – (éto) sportsmen.
doctor is athlete
'The doctor is the athlete'
b. Vrač - *(éto) vrač
doctor is doctor
'The doctor is the doctor.'

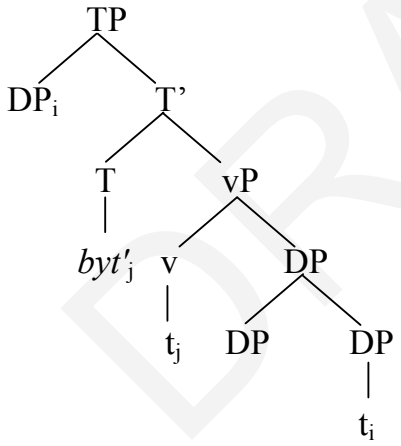
[6] *Byt'* 'be', shown in the trees below, is the non-finite form of the copula *est'* 'is' discussed above.

This structure in (22a) suggests that an entire copular construction will be spelled out in the same phase, and that this is unaffected by the presence or absence of an overt copula. To fit the framework of Distinctness, however, we were lead to a different assumption, namely that an overt copula introduces a phase head, causing the first and second DPs to be spelled out in separate phases, as in (23).

(23) a. Null copula



b. Overt copula



This contrast in structure between overt and null copula constructions calls out for independent support. Unfortunately, vP diagnostics such as those involving vP-modifying adverbs (Beck & Johnson 2004) are difficult to apply to copular constructions. Many other diagnostics are only relevant if the v in question is associated with an event argument (e.g. López-González 2010) or theta-role assignment (e.g. Pereltsvaig 2001), which is not presumably not the case for this semantically bleached item.[7]

[7] If the idea of base-generating the overt copula in v is ultimately untenable, it is useful to bear in mind that this analysis does not require v specifically. For example, the overt copula could alternatively be treated as introducing a small clause CP, which would still allow the two DPs to be linearized separately.

Lest a PF constraint appear too appealing at this point, data like (24) argue in favour of a Distinctness analysis. The two homophonous DPs here are not consecutive, so the ungrammaticality of (24b)[8] cannot be explained by the PF constraint that has been entertained here. Under a Distinctness analysis where the copula introduces a phase head, however, the ungrammaticality of (24b) is expected: regardless of string adjacency, the DPs are linearized together unless a phase head (e.g. an overt copula) intervenes to create a new spell-out domain.

(24) a. Vrač ne est' vrač.
doctor not is doctor
b. *Vrač ne vrač.
doctor not doctor
'A doctor is not a doctor'

Similar support for a Distinctness analysis is provided by coordination of multiple *wh*-phrases as in (25).[9]

(25) Kto i čto zaxvatil? (Gribanova 2009:135)
who and what grabbed
'Who grabbed what?'

Coordination, while destroying string adjacency, does not salvage instances of homophonous *wh*-phrases like (1a), shown in (26).

(26) ??Čto i čto obuslovilo?
what and what conditioned
'What conditioned what?'

The utterance in (26) would not be subject to the PF constraint since the homophonous DPs are not consecutive. If we assume that they are spelled out within the same phase, as was the case for non-coordinated multiple-*wh* questions like (15), Distinctness again captures the ungrammaticality.[10]

[8] Interestingly enough, (24b) is judged grammatical if the first *vrač* 'doctor' is interpreted specifically, showing a contrast similar to the data in (20).

[9] Note that we do not expect (25) to be problematic at linearization, as the first DP is nominative and the second is accusative (though syncretic with the nominative form), which allows them to be distinguished at linearization as the pair <DP_{nom},$DP_{\sout{acc}}$>. Furthermore, as pointed out by a reviewer, the DPs differ in animacy, which Russian may be sensitive to at linearization as well.

[10] Here the unlinearizable pair is <$DP_{\sout{nom}}$,$DP_{\sout{acc}}$>.

3. Some outstanding issues

Before concluding, there are a few relevant issues that have gone undiscussed so far. The first is the role of prosody. Not having performed a systematic investigation, I cannot assert that prosodic factors are not influencing grammaticality in these sentences. For contrasts in (13), however, such an explanation seems unlikely, since the case- and gender-matching non-homophonous DPs (13b) are unlikely to bear different prosody from their non-case-matching and non-gender-matching counterparts (13a,c).

Another issue is the role of tautology. One could imagine that sentences like (4a) are ungrammatical due to their ostensible semantic vacuity. This, however, does not predict the grammaticality of (4b). Furthermore, analyses that tie tautology to ungrammaticality, notably Gajewski (2002), consider tautological constructions those which are true in virtue of their form (e.g. **There is every new student*), which is not the case for the sentences examined here.

Finally, this paper has focused on Russian, but Bošković's PF constraint is intended to apply to a number of multiple-*wh* fronting languages, including Bulgarian, as demonstrated below.

(27) a. *Kakvo kakvo obuslavja? (Bošković 2002:364)
what what conditions
b. Kakvo obuslavja kakvo?
what conditions what
`What conditions what?'

Bulgarian, however, provides data which may be challenging for a Distinctness analysis. The *wh*-phrase *kakvo* 'what' can be phonologically reduced to *kvo*, and if one (but not both) of the *wh*-phrases is reduced, the sentence becomes acceptable to some speakers (Boris Harizanov, p.c.), as shown in (28).

(28) a. Kvo kakvo obuslavja?
what what conditions
b. ?Kakvo kvo obuslavja?
what what conditions
c. *Kvo kvo obuslavja?
what what conditions
'What conditions what?'

A phonological contrast like this is not expected to affect linearization, so Distinctness appears to be an unlikely explanation for this set of data. As brought up when dealing with case syncretism in (15), however, linearization has access to some pseudo-phonological information such that case features are impoverished among items sharing a phonological form. The *wh*-phrases in (28a) and (28b) are no longer case syncretic in quite the way they were in (27) and (28c), and it could be posited that when the surface forms are not sufficiently similar, case features are not impoverished.

A similar problem arises when sluicing data is considered. Richards considers certain prohibitions in multiple sluicing to be due to Distinctness. The sentence in (12a), for example, is ruled out because the *wh*-phrases *who* and *who(m)* are indistinguishable at linearization and are spelled-out in the same phase, violating Distinctness.

This pattern can also be seen in Russian, as shown in (29).

(29) *Vse obuslovilo čto-to, no ja ne pomnju čto čto.
all conditioned something, but I not remember what what
'Everything conditioned something, what I don't remember what conditioned what.'

In Bulgarian, however, such sluicing is not completely ungrammatical, which is difficult to reconcile with the data in (29).

(30) Nešto obuslavja nešto, no ne znam kakvo ?(y) kakvo.
something conditions something but not know what (and) what
'Something conditions something, but I don't know what conditions what.'

The fact that (30) is improved with *wh*-coordination, which destroys string adjacency, supports the idea from (28) that Bulgarian is somehow more sensitive to the surface form of an utterance than Russian is.

4. Conclusion

In Section 1 we saw a ban on consecutive homophonous *wh*-phrases extend to consecutive homophonous noun phrases. This pattern was then re-explained in Section 2 through a general constraint on linearization, Distinctness, which is about neither homophony nor linear adjacency but which can explain the ungrammatical examples above as extreme cases of $<\alpha,\alpha>$.

Throughout I have argued for the use of Distinctness over a PF constraint against consecutive homophonous noun phrases. Though data like (2) and (28) (and perhaps that in footnote 5) fare better under a PF-constraint analysis, a host of examples cannot be explained by this constraint, as is. These include (24) and (26) where the homophonous elements are not consecutive but are still ungrammatical, as well as (6)-(8), (20), (21), and (29) where the homophonous items are consecutive but not ungrammatical.

Overall, Distinctness provides a more general account of the repetition avoidance pointed out by Bošković without requiring the stipulation of PF constraints. Some ostensible problems with this account surfaced, and the solutions proposed by Distinctness were to treat the similar elements as possessing different features that are visible at linearization (e.g. (6)-(7)) or to treat the similar elements as occurring in different phases (e.g. (4)-(5)). Here, that leads us to treat overt and null copula constructions as involving different structures, as in (23). This Distinctness analysis supports theories of Distributed Morphology that treat case syncretism as involving impoverished case features, and it supports treating Russian noun phrases as maximally DPs, not NP (like Pereltsvaig 2007), since noun phrases are only subject to Distinctness if they involve functional heads[11]. Still, a number of the assumptions made here in order to facilitate a Distinctness analyses, such as the structures in (23) and the visibility of rich features at linearization (e.g. those suggested to handle (20)) call out for further research.

Acknowledgements

Many thanks to my informants Dmitriy Krasny, Julia Yarmolinskaya, Alëna Balasanova, Evgueni Oppredelennov, Olga Kharytonava, Boris Harizanov, and Christo Kirov, as well as to the audience at the 2010 Georgetown Linguistics Society conference, Bob Frank, Geraldine Legendre, Kyle Rawlins, and the audience and anonymous reviewers for FDSL 8.5.

References

Beck, Sigrid & Kyle Johnson. 2004. Double Objects Again. *Linguistic Inquiry*. 35, 97-124.

Bonet, Eulalia. 1991. Morphology after Syntax: Pronominal Clitics in Romance Languages. Doctoral dissertation, Massachusetts Institute of Technology.

Bošković, Željko. 2002. On multiple wh-fronting. *Linguistic Inquiry*. 33, 351-383.

Fukazawa, H. 1999. Theoretical implications of OCP effects on features in Optimality Theory. Doctoral dissertation, University of Maryland.

[11] Here D is a functional head, N is not.

Gajewsky, Jon. 2002. On analyticity in natural language. Unpublished manuscript.
Gribanova, Vera. 2009. Structural Adjacency and the Typology of Interrogative Interpretations. *Linguistic Inquiry*. 40, 133-154.
Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Linguistic Inquiry Monograph No 25. Cambridge, MA: The MIT Press.
Legate, Julie Anne. 2003. Some Interface Properties of the Phase. *Linguistic Inquiry*. 34, 506-515.
López-González, Mónica. 2010. Ser and Estar: Their syntax, semantics, pragmatics, and acquisition in Mexican Spanish. Doctoral dissertation, Johns Hopkins University.
Pereltsvaig, Asya. 2001. On the Nature of Intra-Clausal Relations: A Study of Copular Sentences in Russian and Italian. Doctoral dissertation, McGill University.
Pereltsvaig, Asya. On the Universality of DP: A View from Russian. *Studia Linguistica*. 61(1), 59-94.
Richards, Norvin. 2010. *Uttering Trees*. Cambridge, MA: The MIT Press.
Ross, John. 1972. Double-ing. *Linguistic Inquiry*. 3, 61-86.
Walter, Mary Ann. 2007. Repetition Aviodance in Human Language. Doctoral dissertation, Massachusetts Institute of Technology.