

A morphological solution to agreement puzzles in Slavic

AKIRA WATANABE

University of Tokyo

akirawat@l.u-tokyo.ac.jp

1. Introduction

In this paper, I propose that underspecification in narrow syntax can be restored to full specification in morphology. Restoration to full specification in morphology gives rise to what one may call disguised syntactic underspecification, in the sense that underspecification is not transparently visible on the PF side.

Syntactic underspecification is motivated by the predictability of feature values. I would like to suggest that in each case, underspecification is keyed to one particular kind of predictors of feature values as a parametric option. On the morphological side, I claim that missing predictable feature values are supplied in morphology only when the predictor of the value is a feature available in morphology. Thus, the type of the value predictor serves as a diagnosis of disguised syntactic underspecification.

The phenomenon to be taken up is failure of agreement with subjects that have a higher numeral in a subset of Slavic languages including Czech, Polish, Serbo-Croatian, and Slovene (see Veselovská [1] for Czech, Rappaport [2] and Rutkowski and Maliszewska [3] for Polish, Franks [4] and Wechsler and Zlatić [5] for Serbo-Croatian, and Marušič and Nevins [6] for Slovene; see also Corbett [7] and Franks [8]). A Czech example is given in (1) to illustrate the point.

- (1) Těch pět hezkých dívek jelo. Czech
this.gen.pl five beautiful.gen.pl girl.gen.pl traveled.n.sg (Franks [8], p. 137)

Here, the neuter singular form of the verb is obligatory, despite the fact that the head noun takes the (genitive) plural form. It is a mistake to attribute the lack of plural agreement directly to the genitive marking of the head noun, since Russian allows plural agreement despite genitive marking, as in (2).¹

- (2) Pjat' krasivyx devušek prišli. Russian
five beautiful girl.gen.pl arrived.pl (Franks [8], p. 106)

¹ Neuter singular is also possible in Russian, analysis of which is beyond the scope of this paper. See Bošković [9], Franks [4, 8], Pereltsvaig [10], and the references cited there for various possibilities. See also Corbett's [7] p. 215 quantitative corpus data, which exhibits a huge difference between Russian and the set of Slavic languages that concern us here.

Let me also note that care must be taken to exclude from consideration subjects made exceedingly long by modifiers like relative clauses, since such examples will invite the left dislocation parse with the null subject actually triggering plural agreement. Significantly, West and South Slavic languages allow null subjects (Franks [8]). See Sturgeon [11] for left dislocation in Czech.

- (6) a. Pet devojaka i nekolikomomaka je skakalo. Serbo-Croatian
 five girl.gen.pl and several boy.gen.pl aux.3sg jumped.n.sg
 b. Tele i dete su skakali. Serbo-Croatian
 calf.n.sg and child.n.sg aux.3pl jumped.m.pl (Franks [8], p. 115)

If feature resolution under coordination is based on specification in syntax, coordination of unspecified [singular] can only yield unspecified [singular], a correct prediction⁵.

Though underspecification of [singular] in narrow syntax can explain why the presence of a higher numeral forces the default neuter singular agreement, it makes the genitive plural form of the head noun puzzling. Where does plural marking come from if [singular] is valueless? To solve this problem, we need to take into account the number feature system as a whole. I adopt the system in (7), which employs [\pm augmented] in addition to [\pm singular] to define the basic number distinction (see Harbour [15] and the references cited there).

(7) The Number Feature System

- i. singular: [+singular, -augmented]
 ii. dual: [-singular, -augmented]
 iii. plural: [-singular, +augmented] (3)

Significantly, [+singular, +augmented] is an impossible combination under this system. This means that the minus value of [singular] in the feature makeup of plural is predictable given [+augmented]. And this is the conceptual basis for syntactic underspecification, which feeds semantic interpretation: predictable values can be omitted. Significantly, with numerals 5 and above, the number features must be specified as [-singular, +augmented], if full specification is used.

Let us now suppose that the predictable value of [singular], though absent in narrow syntax, is supplied in morphological computation. This value insertion can be implemented by (8).

- (8) [singular] \rightarrow [-singular] / [_____, +augmented]

The background assumption here is that morphological computation is allowed to operate on the output of narrow syntax computation before morphosyntactic features receive phonological realization, as in Distributed Morphology (Halle and Marantz [16]).

One might wonder whether [\pm augmented] is motivated for languages like Czech, Polish, and Serbo-Croatian which lack the category of dual. Slovene preserves dual, but it was lost in Czech, Polish, and Serbo-Croatian. At this point, we can appeal to Watanabe's [17] proposal that [\pm augmented] is involved in the licensing of numerals. Specifically, he suggests (9) as a universal principle.

- (9) Numerals are licensed only when the # head is marked for [\pm augmented].

⁵ For further intricacies of the agreement pattern triggered by coordinated subjects, see Bošković [14] on Serbo-Croatian and Marušič & Nevins [6] on Slovene.

According to this hypothesis, [\pm augmented] is active in those languages which have numerals, even if dual is absent as an inflectional category. It follows that the use of [\pm augmented] is legitimate for languages like Czech, Polish, and Serbo-Croatian.

Incidentally, Watanabe [17] points out that the system in (7) allows [augmented] to be unspecified in the presence of [+singular], since the minus value is predictable in that case. Underspecification of [augmented], then, blocks the appearance of numeral 1, given (9). This phenomenon is fairly common with numerical bases, as documented by Hurford [18, 19]. Thus, the incompatibility of [+singular, +augmented] leads to two logically possible types of underspecification in the number feature system. For more details about the relation between [augmented] and numerals, see Watanabe [17].

The discussion so far has taken care of neuter singular subject-verb agreement and plural marking of the head noun. One remaining piece in the picture is genitive case. For this, I simply assume syncretism, of the form in (10i).

- (10) i. structural case (nom, acc) \rightarrow gen / _____ [(+augmented), -less.than.a.handful]
 ii. oblique \rightarrow no change

(10i) is nothing more than a descriptive statement, and I leave vigorous formalization to future research. It will require an in-depth analysis of the entire case system in Slavic and more generally in Universal Grammar. Let me just mention that nominative, accusative, and genitive are implicated in other types of case syncretism in Slavic (see various chapters in Comrie and Corbett [20]), so that something like (10i) is a very plausible additional candidate.

(10ii) is intended to capture the fact that in Czech, Polish, and Slovene, both the numeral and the head noun exhibit the expected case form in oblique positions, as illustrated in (11) for Czech⁶.

- (11) s pěti pány Czech
 with five.inst man.inst.pl (Franks [8], p. 136)

It is worth reiterating at this point that it is not true that genitive marking of the head noun blocks agreement. In Polish predicate adjective constructions, the adjective itself appears in the genitive plural form, as in (12), despite the fact that the copular verb is neuter singular.

- (12) Szesc kobiet bylo smutnych. Polish
 six-nom woman.gen.pl be.past.n.sg sad.gen.pl (Dziwirek [21], p. 147)

This means that the subject with a numeral is in principle capable of entering into the agreement relation. A peculiarity of Polish is that adjectives come with [\pm augmented] and [\pm less.than.a.handful] (in addition to [\pm singular]) as inflectional features, agreeing with the subject with respect to these two features. (8) and (10i) then ensure that the predicate adjective will take the genitive plural form. Verbs, on the other hand, lack

⁶ In Serbo-Croatian, the head noun appears in the genitive plural form in oblique contexts, too, as discussed in Franks [4, 8] and Wechsler and Zlatić [5]. This difference seems to be correlated with the fact that the numeral is invariant in form irrespective of case in Serbo-Croatian, unlike in Czech, Polish, and Slovene. I will return to case forms of numerals below.

[±augmented] and [±less.than.a.handful] as well as case, preventing (8) and (10i) from applying⁷.

The agreement in terms of [±augmented] and [±less.than.a.handful] is not limited to predicative adjectives in Polish. In the Czech example in (1), repeated here, the demonstrative and the attributive adjective take the form of genitive plural.

- (1) Těch pět hezkých dívek jelo. Czech
 this.gen.pl five beautiful.gen.pl girl.gen.pl traveled.n.sg (Franks [8], p. 137)

This morphological shape also arises through (8) and (10i), which in turn are dependent on the working of agreement with respect to [±augmented] and [±less.than.a.handful].

3. Further consequences

As mentioned above, the default neuter singular subject-verb agreement is triggered by numerals 5 and above. The paucal numerals 2, 3, and 4 behave differently. The proposal outlined in the previous section enables us to understand why such a contrast exists.

Important for the purposes of this paper is the fact that the finite verb shows plural agreement when the subject contains a paucal numeral, as illustrated in (13) for Serbo-Croatian.

- (13) a. Dva srpska glumca su otišla / otišli. Serbo-Croatian
 two Serbian actor.m.234 aux.3pl left.m.234/left.m.pl
 b. Dve srpske glumice su otišle Serbo-Croatian
 two Serbian actress.f.234 aux.3pl left.f.pl(=f.234)

(Wechsler & Zlatić [5], p. 151)

Czech, Polish, and Slovene pattern in essentially the same way, except that Slovene uses dual for 2 (Corbett [7]). In other words, there is no syntactic underspecification for the paucal numerals. Higher numerals, on the other hand, force neuter singular, as shown in (14)⁸.

- (14) Pet ljudi je došlo na miting. Serbo-Croatian
 Five people.gen.pl aux.3sg arrived.n.sg at meeting (Franks [8], p. 116)

This contrast between the paucal numeral and the higher numeral has not received a satisfactory account in the past. Franks [8] (p. 128, note 29) suggests that the adjectival status of the paucal numerals is responsible for plural agreement.⁹ Though it is true that the Serbo-Croatian paucal numerals have three distinct case forms (nominative-accusative, genitive, oblique) unlike higher numerals, which are invariant, and that *dva*

⁷ Given the existence of dual in Slovene verbs, further contextual conditions must be added to (8) and (10i), mentioning categorial information about applicable domains, to ensure the singular verb agreement. According to Corbett [22] (p. 134) and Marušič and Nevins [6] (note 1), the Slovene predicative adjective behaves in the same way as the Polish counterpart.

⁸ Serbo-Croatian marginally allows plural, too (Franks [4, 8]; Wechsler and Zlatić [5]). I put this pattern aside, as essentially belonging to a different grammar.

⁹ Veselovská [1] seems to follow suit.

‘two’ in addition is gender-sensitive, the account does not generalize to other languages¹⁰. Czech, for example, distinguishes two case forms for 5 and four case forms for 3, as shown in (15).

(15) Inflectional forms of Czech numerals (Veselovská [1])

a. five: pět (nom, acc), pět-I (gen, dat, loc, instr)

b. three: tři-I (nom, acc, gen), tři-EM (dat), tři-ECH (loc), tři-MI (instr)

There is no sense in which three, but not five, is adjectival in nature. Besides, even if the adjectival status of the paucal numerals turns out to be relevant, it remains to be explained why the paucal numerals are adjectival, unlike higher numerals. The suggestion is nothing more than a restatement of the original problem¹¹.

Corbett [7], on the other hand, speculates that groups with a large cardinality “*are less individuated and are conversely more likely to be viewed as a unit*” (p. 217), leading to the nominal conception of the numeral in question and facilitating singular agreement. To the extent that the categorial status is invoked, Corbett’s proposal runs into the same problems as Franks’. Furthermore, the coordination data in (6) refutes the idea that singular agreement is triggered by the [+singular] feature. The failure of plural agreement in (6a) suggests that it is a mistake to attribute singular agreement forced by the subject with a higher numeral to the semantic notion of singularity, since coordination of semantically meaningful singulars yields plural agreement as in (6b). What is triggered by higher numerals is nothing more than default agreement.

There is further evidence that cardinality itself should not be blamed. Serbo-Croatian has two versions of *many*, one of them behaving like higher numerals and the other triggering plural agreement, as shown in (16).

- (16) a. Mnogo srpskih pisaca je otišlo. Serbo-Croatian
 many Serbian.gen.pl writer.gen.pl aux.3sg left.n.sg
- b. Mnogi mladići su protestovali. Serbo-Croatian
 many.nom.m.pl young.man.nom.m.pl aux.3pl protested.m.pl

(Wechsler and Zlatić [5], pp. 116, 118)

As far as I am aware, there is no cardinality difference between the two versions of *many*. Plural agreement in (16b) shows that a large cardinality does not necessarily trigger singular agreement in this language. The idea of (non-)individuation is mistaken, too. Franks [4] (p. 626) [8] (p. 116) observes that both group and individuated readings are available for (14). Thus, it is an error to associate the idea of less individuated group members with a large cardinality to account for the pattern of subject-verb agreement. What is going on, then?

¹⁰ Wechsler and Zlatić [5] (p. 149) remark that the genitive and oblique forms are in fact not used with 3 and 4, which are essentially frozen. This denies the adjectival nature of these two numerals.

¹¹ It should also be noted that Polish higher numerals change their shape in agreement with the gender information of the head noun in structural case contexts. Their inflection is therefore “adjectival”. See Franks [4, 8] and Rappaport [2] for discussion.

I claim that underspecification holds the key, again. An important auxiliary hypothesis is that syntactic underspecification is enforced in a set of contexts provided independently. In Slavic, [\pm less.than.a.handful] divides plural numerals into two classes. Significantly, this division affects the case form of the head noun, in addition to the value of [\pm singular], the target of underspecification. In Serbo-Croatian, the paucal numerals force the head noun to take what Browne [23] calls the 234 form, which is syncretic with genitive singular for masculine and neuter nouns and with nominative plural for feminine nouns (Franks [8], p. 125, note 8). Higher numerals, on the other hand, require genitive plural, as discussed above. The relevant forms of a masculine noun are given in (17).

- (17) a. prózori (nom.pl) ‘window’ Serbo-Croatian
 b. prózora (gen.pl)
 c. prózora (234 = gen.sg) (Browne [22], p. 319)

Thus, the division would be needed even if these languages did not choose underspecification of [\pm singular]¹².

A similar classification of numerals is found in Irish, quite independently of underspecification. The phenomenon sensitive to the division is initial mutation affecting the head noun that follows the numeral. The dividing line here is between 6 and 7. Lenition applies in (18a), while we find nasalization in (18b).

- (18) a. dhá/trí/ceithre/cúig/sé chat (lenition) Irish
 two/three/four/five/six cat.sg
 b. seacht/ocht/naoi/deich gcat (nasalization)
 seven/eight/nine/ten cat.sg (Acquaviva [24], pp. 165, 167)

The head noun appears in the singular form in both cases, and thus does not vary in number marking. See Acquaviva [24] and the references cited there for further discussion of Irish numerals.

We can conclude that the numeral system can introduce within itself a division that affects the morphosyntax of the head noun in various ways as a parametric option. In the relevant Slavic languages, I claim that case is the primary factor in this division, as encoded by whatever is the correct analysis of syncretism described by (10i). Underspecification of [\pm singular] chooses a subclass of numerals already given in terms of case considerations. It then follows that the subclass in question must be [$-$ less.than.a.handful], since syntactic underspecification is based on predictability of the unspecified value. Crucially, in the class defined by [$+$ less.than.a.handful], the value of [singular] is not always predictable from the value of [augmented]. 2 is specified as [$-$ singular, $-$ augmented], but [$-$ singular] is not predicted by [$-$ augmented], which is

¹² In Czech and Polish, the head noun after the paucal numerals appears in the form required by an external case assigner. Hence nominative plural in the subject position, though with some twists for masculine human nouns in Polish (Rappaport [2]). Genitive plural after higher numerals, on the other hand, is a common Slavic trait. The division thus affects case forms in languages other than Serbo-Croatian as well.

compatible with [+singular] as well, as far as combinatorial possibilities of feature values are concerned. Therefore, the class of [+less.than.a.handful] cannot be chosen as the domain of underspecification.

Note that this result hinges on the assumption that the predictability of the [\pm singular] value comes from [+augmented], but not from the numeral itself. If the numeral itself acts as the predictor of the unspecified value, underspecification can be enforced anywhere. And this latter possibility is found in languages like Hungarian, where any numeral forces the head noun to be in the singular. We then have the following typology of syntactic underspecification of [\pm singular] in the context of the numeral:¹³

(19) Underspecification of [\pm singular] in the context of the numeral

	predictor	domain
i. Hungarian Type	numeral	all
ii. Slavic Type	[+augmented]	[-less.than.a.handful]

These two must be the basic types. When there is a numeral, the value of [singular] can be predicted either by [+augmented] or by the meaning of the numeral itself. No other possibility exists. If [+augmented] is the predictor, the entire numeral domain cannot be selected, since singular and dual are [-augmented]. There must be a class of higher numerals independently given by other morpho-syntactic considerations. If the meaning of the numeral itself matters, the entire domain can be chosen. An open empirical question is whether a subclass of numerals can also be picked out as the domain of underspecification, arbitrarily this time, when the numeral acts as the predictor of the [\pm singular] value. So far, no such case is reported. If this third type does not exist, we can say that selection of the domain of underspecification is maximal.

There is another point to be made. As proposed above, the Slavic underspecification is accompanied by value insertion during morphological computation. I suggest that this value insertion is possible because the predictor of the value is [+augmented], a feature available to morphological operations. In other words, the missing value can be supplied in morphology only when that value is predicted by another feature available in morphology.¹⁴ On the other hand, the semantic content of the numeral cannot play a role in morphology. It is simply sent to the LF interface for semantic interpretation. Thus, it is predicted that the Hungarian type underspecification will disallow value insertion in morphology, hence always visible in a transparent way.

4. Comparison with previous analyses

¹³ I put aside vague quantifiers like *many* here, though they also trigger underspecification in Slavic, as we have seen in (16). They belong to a separate system of quantification. See Watanabe [17] for arguments that they are structurally different from numerals.

¹⁴ In order to make this idea work, one probably has to say that the predictor feature (value) is marked as such, hooked to (8), so that its predictor status is visible to the morphological component. I leave it to future research to explore implications of this mechanism.

Let us take stock. A subset of Slavic languages underspecify [singular], based on the fact that [+augmented] must be combined with [-singular]. The underspecification is associated with the # head that selects a class of numerals 5 and above, a division which must be independently provided for case morphology. This association is forced by the choice of [+augmented] as the predictor of the missing value of [singular]. The underspecification in narrow syntax, however, is masked by “repair” in morphology.

Above, we have already seen that the previous analyses have not succeeded in explaining why singular agreement is forced by higher numerals, but not by the paucal numerals. Let us now focus on the part of the accounts that deals with the forced default singular agreement itself in the case of higher numerals.

Franks [8] claims that the plural feature is blocked from percolating to the top projection by oblique genitive marking and hence inaccessible for agreement in Serbo-Croatian, whereas higher numerals are always in the accusative in Czech and Polish, failing to induce agreement. It is highly problematic that a rather intricate common set of agreement problems does not receive a unifying account. Furthermore, we have seen that subject-predicate agreement is not completely blocked, as evidenced by Polish data. Recall that Polish predicate adjectives are inflected as genitive plural in (12). This fact can be accounted for by (8) and (10i) if adjectives have [+augmented] and [-less.than.a.handful]. The value of these features must come from agreement with the subject.

Wechsler and Zlatic [5] propose for Serbo-Croatian that higher numerals lack phi-features and therefore lead to default agreement¹⁵. Polish data on predicate adjectives, again, indicate that default agreement is used for [singular], but not for [augmented].

5. Conclusion

In this paper, I have explored the idea that underspecification in narrow syntax is responsible for singular subject-verb agreement induced by numerals 5 and above in Czech, Polish, Serbo-Croatian, and Slovene. This is an instance of disguised underspecification, since morphology repairs the underspecification of [singular] in narrow syntax. A novel result is that we now have a principled explanation of why the singular agreement is forced by higher numerals but not by the paucal numerals. This explanation is made possible by the very nature of underspecification, which must be motivated by the predictability of the missing feature value. The predictor of the missing feature value is parametrized. In the relevant Slavic languages, it is keyed to [+augmented].

What masks syntactic underspecification is a morphological operation that provides the predictable feature value missing in narrow syntax. This morphological operation must be distinguished from the insertion of the unmarked value advocated by Harbour [25] and Noyer [26]. Quite generally, disguised underspecification can be regarded as a

¹⁵ Veselovská [1], p. 302 seems to assume something similar for Polish.

major source of agreement mismatch. It is an interesting task for future research to apply the idea to various individual cases.

References

- [1] L. Veselovská, Agreement Patterns of Czech Group Nouns and Quantifiers, in *Semi-Lexical Categories*, eds. N. Corver and H. van Riemsdijk, Mouton de Gruyter, Berlin, 2001, pp. 273–320.
- [2] G. Rappapor, Case Syncretism, Features, and the Morphosyntax of Polish Numeral Phrases, in *Generative linguistics in Poland 5*, 2003 pp. 123–137.
- [3] P. Rutkowski and H. Maliszewska. On Prepositional Phrases inside Numeral Expressions in Polish, *Lingua* 117, 2007, pp. 784–813.
- [4] S. Franks, Parametric Properties of Numeral Phrases in Slavic, *Natural Language and Linguistic Theory* 12, 1994, pp. 597–674
- [5] Wechsler, S. and L. Zlatić, *The Many Faces of Agreement*, CSLI Publications, Stanford, 2003.
- [6] F. Marušić and A. Nevins, Two Types of Neuter: Closest-Conjunct Agreement in the Presence of ‘5 and Ups’, paper presented at *Formal Approaches to Slavic Linguistics 18*, 2009.
- [7] Corbett, G. G., *Number*, Cambridge University Press, Cambridge, 2000.
- [8] Franks, S., *Parameters of Slavic Morphosyntax*, Oxford University Press, Oxford, 1995.
- [9] Ž. Bošković, Case and Agreement with Genitive of Quantification in Russian, in *Agreement System*, ed. C. Boeckx, John Benjamins, Amsterdam, 2006, pp. 99–120.
- [10] A. Pereltsvaig, Small Nominals. *Natural Language and Linguistic Theory* 24, 2006, pp. 433–500.
- [11] Sturgeon, A., *The Left Periphery: The Interaction of Syntax, Pragmatics and Prosody in Czech*, John Benjamins, Amsterdam, 2008.
- [12] D. F. Farkas, The Unmarked Determiner, in *Non-Definiteness and Plurality*, eds. S. Vogeleer and L. Tasmowski, John Benjamins, Amsterdam, 2006, pp. 81–105.
- [13] A. Ortmann, Where Plural Refuses to Agree: Feature Unification and Morphological Economy, *Acta Linguistica Hungarica* 47, 2000, pp. 249–288.
- [14] Ž. Bošković, Unifying First and Last Conjunct Agreement, *Natural Language and Linguistic Theory* 27, 2009, pp. 455–496.
- [15] Harbour, D., *Morphosemantic Number*, Springer, Dordrecht, 2007.
- [16] M. Halle and A. Marantz, Distributed Morphology and Pieces of Inflection, in *The View from Building 20*, eds. K. Hale and S. J. Keyser, MIT Press, Cambridge, 1993, p. 111–176.
- [17] A. Watanabe, Vague Quantity, Numerals, and Natural Numbers, *Syntax* 13, 2010, p 37-77.
- [18] Hurford, J. R., *Language and Number*, Basil Blackwell, Oxford, 1987.
- [19] J. R. Hurford, The Interaction between Numerals and Nouns, in *Noun Phrase Structure in the Languages of Europe*, ed. F. Plank, Mouton de Gruyter, Berlin, 2003, pp. 561–620.
- [20] Comrie, B., and G. G. Corbett, eds., *The Slavonic Languages*, Routledge, London, 1993.
- [21] K. Dziwirek, Default Agreement in Polish, in *Grammatical Relations*, eds. K. Dziwirek, P. Farrel, and E. Majías-Bikandi, CSLI Publications, Stanford, 1990, pp. 147–161.
- [22] Corbett, G. G., *Agreement*, Cambridge University Press, Cambridge, 2006.
- [23] W. Browne, Serbo-Croat, in *The Slavonic Languages*, eds. B. Comrie and G. G. Corbett, Routledge, London, 1993, pp. 306–387.
- [24] Acquaviva, P., *Lexical Plurals*, Oxford University Press, Oxford, 2008.
- [25] D. Harbour, The Kiowa Case for Feature Insertion, *Natural Language and Linguistic Theory* 21, 2003, pp. 543–578

- [26] R. Noyer, Impoverishment Theory and Morphosyntactic Markedness, in *Morphology and its relation to phonology and syntax*, eds. S. G. Lapointe, D. K. Brentari, and P. M. Farrell, CSLI Publications, Stanford, 1998, pp. 264–285.