Czech is not [S[VO]] – A reply to Šimík & Jasinskaja

Hubert Haider♣ & Luka Szucsich♠

♣Dept. of Linguistics & Centre for Cognitive Neuroscience, Univ. Salzburg – ♠Slavic Dept., HUB Berlin.

In the open-review commentary to our target paper[1], posted on Lingbuzz (006312), Šimík & Jasinskaja (Š&J) put a most welcome focus on Czech. They subscribe to the widely entertained view according to which Czech is a language with a basic clause structure much like English, that is, an SVO language, as they explicate in Jasinskaja & Šimík (forthc.). We are grateful for their thought-provoking attempts of challenging our theory thoroughly by confronting it with properties of a Slavic language we have considered only in passing in our target paper. The essential quality test for a theory is thorough falsification trials. We gladly grasp the opportunity to demonstrate that our theory stands the test successfully.

Š&J find fault with four properties we attributed to Slavic languages, when applied to Czech, namely (ii) (vi), and (viii), plus the property (ix), which they add although it is a subcase of (vii.). For ease of reference, we juxtapose their summary table and the version we shall defend.

ambidir.	Russian	Czech
yes	yes	yes
no	yes	yes
no	no	no
no	no	no
no	yes	yes
yes	yes	no
no	no	no
no	no	yes
yes?	yes	no

Syntactic properties		[S[VO]]	ambidirect.	RUSSIAN	CZECH
i.	S-V-O as an acceptable order	yes	yes	yes	yes
ii.	obligatory preverbal subject	yes	no	no	no
iii.	subject wh-in-situ restriction	yes	no	no	no
iv.	adverbial wh-in-situ restriction	yes	no	no	no
v.	LLC for left-adjoined adjuncts	yes	no	no	no
vi.	fillers for left branch gaps	no	yes	yes	yes
vii.	rigid word order	yes	no	no	no
viii.	rigid relative order of auxiliaries	yes	no	no	yes
ix.	free OV/VO word order	no	yes	yes	yes

Table 1   Š&J's summary   H&S' summary

Let us start with (viii) and an 'operating instruction' for the list of syntactic properties in the table above. It is important to keep in mind that the values in the [S[VO]] column are values of *necessary* properties of [S[VO]] languages. In other words, if the grammar of a language does not meet the respective property for "yes" or for "no", the respective language cannot be an [S[VO]] language. Crucially, the properties are not *sufficient* ones, that is, even if a language meets one or the other of the criteria, it need not be an [S[VO]] language. This is important for the evaluation of (viii). We gratefully accept the information that in Czech, there are invariant orders of auxiliaries.[2] This does not make Czech an SVO language, however. The same is true for Frisian, a Germanic SOV language without variation of auxiliary verb order, and many other SOV languages.

---

[1] Haider, H. & L. Szucsich. (in press, to appear 2022). Slavic languages – "SVO" languages without SVO qualities? *Theoretical Linguistics*. [*lingbuzz/*004973].

[2] Š&J's Czech example does not match the Polish counterpart. As the authors concede, it involves a semi-lexical verb *dát* 'give' and a modal *chtít* 'want', which systematically allows for finite embeddings with the subjunctive marker *by*, introduced by a complementizer *aby* 'that'. In contrast, the Polish modal *musieć* 'must' does not show this behaviour, and B/C/S *morati* 'must' only selects finite complements in varieties which have lost the infinitive or, at least, pushed it back. So, syntactically, Polish *musieć* (and Czech *muset*) differ from *chcieć* (and *chtít*).

Frisian is exceptional in this respect since other Germanic SOV languages display at least some variable orders, but in East Asian SOV languages, the order is invariable, too. Note however, that this does not falsify our claim. All we claim is that a language cannot be an [S[VO]] language if the relative order of auxiliaries is variable. So, if there is no variation in Czech, we conclude that Czech is the Slavic counterpart of Frisian in this respect, as a language that does not make use of a principally available and compatible syntactic option of the respective type of the language (family).

Let us continue with the objection concerning line (ix), which Š&J added to our table. In fact, this is a sub-instance of (vii.). Here, "free" has to be interpreted as *syntactically* free, that is, not restricted by *syntactical* constraints, or as Siewierska & *Uhliřová* (2010:109) put it: *"In each of the Slavic languages, all twenty-four possible combinations of a subject, direct object, indirect object and verb occur as grammatical declarative orders."*

In SOV or SVO languages, nominal objects are syntactically restricted by the directionality requirement of the head they depend on. Nominal objects *obligatorily* precede the base position of the verb in SOV. In SVO, they *obligatorily* follow the verb. The order restriction holds independently of information structure properties, intonation effects, or other pragmatic preferences. In Czech, and in fact in all Slavic languages, even in Sorbian to a minor extent,[3] objects as well as the subject may precede or follow the verb. This fact cannot be seriously contested.

However and crucially, "free" must not be interpreted as "anything goes anytime". The existence of a pragmatically unmarked or neutral order is fully consistent with the syntactical word order potential. The fact that alternatively available orders are not always freely exchangeable in utterances is an independent issue, however. Information structure preferences partition the pool of variants. We do not want to repeat the explication of the interface effects presented in Haider (2020). Let us merely summarize it: "*In general, when syntax admits structural variation, this potential is captured and utilized by other subsystems of grammar.*" (Haider 2020: 375). All variants are grammatical, but, of course, they are not equivalent with respect to information structuring since they may be associated with particular focus, topic, or givenness properties. Importantly, the conditions of information structuring do not *constrain* syntax; they merely *exploit* the syntactically available options.[4] On the other hand, if syntax does not admit variation, information structuring cannot coerce syntax, otherwise all languages would closely resemble Czech or Russian.

It is a fact of Czech that there are acceptable utterances in which a direct object precedes a main verb and that there are acceptable utterances in which a direct object follows a main verb, finite or not, in main clauses as well as in embedded ones. The relevant data are familiar and have been reconfirmed on independent grounds, for instance by computational methods, measuring the word order freedom in parsed corpora (tree banks). Kuboň (et als. 2016) present the following percentages for Czech, on the basis of a tree-bank corpus of 16,862 matching sentences out of 87.913 *main clauses* in total, in comparison with 22 other languages (1).

---

[3] Even Sorbian – although predominantly verb-final – is not *strictly* SOV, as (i) illustrates (Scholze 2015: 206):
   (i) Ćeta *dari* mi rjaneho žurka (aunt *gave* me beautiful hamster).
[4] "*Another significant property of Slavic languages is their relatively free word order, which generally serves to express functional sentence perspective information rather than grammatical relations.*" Franks (2005: 376).

(1) Czech: SVO: 51.2% , VSO: 9.6%, VOS: 10.0%; OVS: 21.4%, SOV: 5.8%, OSV: 2.0%.

The percentages contrast sharply with undisputed SVO languages such as English or Portuguese, for instance. (2) and (3) list the percentages of SVO vs. SVO serializations in main and embedded clauses, respectively. Equally telling is the high percentage of the residual orders ("rest") in Czech. This is the hallmark of "free word order" languages. The percentages of Czech very closely resemble the percentages of Slovak and Slovenian.

(2)   main clause orders:
      a. Czech: SVO 51,2% vs. SOV 5,8% orders;   **rest: 43%**
      c. English: SVO 83,1% vs. SOV 0,0% orders total;   rest: 17%
      d. Portuguese: SVO 80,7% vs. SOV 3,1% orders total;   rest: 16%

(3)   embedded clauses:
      a. Czech: SVO 60,2% vs. SOV 10,4% orders;   **rest: 29%**
      c. English: SVO 96,9% vs. SOV 0,0% orders total;   rest: 3%
      d. Portuguese: SVO 76,0% vs. SOV 11,8% orders total;   rest: 12%

The relatively low percentage of SVO orders and the high percentage of residual orders (viz. 43% and 29%) mark the crucial difference between a T3 language such as Czech and a genuine SVO language such as English. Kuboň et als. do not differentiate between pronominal and non-pronominal objects. This accounts for "SOV" orders in VO languages such as Portuguese, with cliticized pronominals.

Kuboň (et als. 2016:15) computationally compare and then rank their set of 23 languages by computing four measures of word order variation: max-min, Euclidian distance, cosine similarity, and entropy. On each of the four measures, Czech ends up in the top group of five languages with respect to word order freedom, namely Ancient Greek, Czech, Latin, Slovak and Slovenian. SVO languages, such as English and Portuguese, are in the opposite region, namely in the top group of languages with highly restricted word order. This is independent evidence for an essential difference between Czech and its kin languages on the one hand, and uncontroversial [S[VO]] languages on the other hand, and it supports the claim put forth in the paper. Merlo & Samo (this vol.) confirm this result by measuring the entropy difference between SVO languages and Slavic.

Let us turn now to the next trait, namely the *obligatory preverbal subject position* of SVO languages. Actually, this is a type defining property. The [S[VO]] clause structure implicates an obligatory structural position for the syntactic subject of the clause. This position is outside of, and preceding, the VP. An [S[VO]] clause is ungrammatical when this position is radically empty. This is the case when there is no subject argument available and the position is not filled with an expletive, as in the French example (4a), contrasting with (4b):

(4) a.*Dimanche 24 mai, a été procédé à l'installation du conseil municipal.
      Sunday 24 may, has been proceeded to the-installation of-the council municipal
   b. Dimanche 24 mai, *il* a été procédé à l'installation du conseil municipal.[5]

The minimal instance of a subjectless construction is the passive of intransitives. If an intransitive verb is passivized, there is no argument left for the subject position and so it must be filled

---

[5] https://www.haut-bocage.fr/2020/election-du-maire-et-des-adjoints/   [Nov. 28, 2021]

with an expletive (4b) in SVO. The expletive is a 3rd person sg. pronoun in French. Such a pronoun falls prey to pro-drop in Romance pro-drop languages. Consequently, Romance pro-drop languages cannot and do not passivize intransitive verbs since the expletive must not be a null pronoun[6] (Haider 2019). The same is true for English, for a different reason though. English lacks a suitable expletive since "*it*" as well as "*there*" turn out to be inept (see Haider 2019).

The direct test for a syntactically obligatory subject position is the passivization of intransitive verbs because in this case, there is no object that could be promoted to a derived subject. In SVO, in contrast to T3, SOV and VSO, the resulting clause structure is ungrammatical without a subject expletive. Czech provides such constructions. They are grammatical and fully acceptable in spite of the absence of an expletive. Consequently, Czech cannot be [S[VO]].

(5)   a. Bylo pracováno s celkovými koncentracemi.[7]
          was worked with total concentration
      b. V tomto výzkumu *bylo pracováno* s konceptem statistické významnosti.
           in this research was worked with concept (of) statistical significance
      c. Rozkazubylo uposlechnuto.
          *the order*Dat. *was obeyed*

Š&J present several unacceptable examples of intransitive passive examples from Czech and Russian and conclude "*Czech – and this time also Russian – pattern with Italian and Spanish in this respect.*" However, this is exactly not what their data show. In Italian and Spanish, the passive of *any* intransitive verb is ungrammatical. This includes cases such as (5a-c) and many others. However, (5a-c) are acceptable and grammatical in Czech. Š&J argue as if we had claimed that *any* passivized intransitive verb is fully acceptable in Czech or Russian. This we did not claim and we would not claim, simply because it is wrong, not only for Czech.

What we claim is this: If an intransitive verb can be passivized in an [S[VO]] language, the subject position must not remain empty. If it is empty, the result is ungrammatical, as in English, unless an expletive is adduced for filling the obligatory subject position, as in French. In SOV and T3 languages however, there is no obligatory structural subject position, hence no room for a subject expletive. Czech behaves as expected and predicted for a T3 language, and so do other Slavic languages.

The passive of intransitives is not the only source of evidence, of course, but it is the most syntactically straightforward one. There are numerous papers on Czech impersonal constructions (cf. Guiraud-Weber & Kor Chahine 2013), but data such as (6) are more subtle to analyze properly (see Szucsich 2006). As discussed in detail in Haider (2019:20), unlike expletives, semantically empty subject arguments are licit null-subjects in pro-drop languages. The presence of an accusative object in (6) is an indication of the presence of a null subject. So, technically,

---

[6] The Generative literature is in the wrong in this respect. It fails to appreciate that "empty expletives" would be a corollary for Romance pro drop languages. However, intransitive passives are ungrammatical in these languages exactly because empty expletives are theoretical entities that do not exist in the linguistic reality. See Haider (2019) for details.

[7] The German versions are fully parallel to their Czech counterparts.
  i.   Gearbeiṭet wurde mit höchsten Konzentrationen.   (= 5a)
  ii. Dem BefehlDat. wurde gehorcht.   (=5c)

(6a,b) are not subjectless. In the German counterpart (6c), the semantically empty subject "es" (it) is audible.

(6) a. Bratra zabilo.   [Guiraud-Weber & Kor Chahine (2013:12)]
        brother_{acc} killed_{neut}   ['(Somebody/something) killed my brother.']
     b. Souseda ranilo.
        neighbor_{acc} injured_{neut}
        'The neighbor was injured.'
     c. Plötzlich hat *es* ihn_{Acc} ohne erkennbaren Grund umgeworfen.
         Suddenly has *it* him without noticeable cause knocked-over
         'He suddenly knocked over, without noticeable reason'

What is remarkable, nevertheless, is the position of the accusative object in (6a,b). Preverbal accusative objects are ungrammatical in prototypical [S[VO]] languages, unless they are wh-moved to the clause initial position, which is a possible source of (6). Let us therefore finish the data review for subject properties with another clear-cut set of positive evidence, taken from Guiraud-Weber & Kor Chahine (2013:9). Czech and German share the very same construction, namely a copula construction with dative plus a PP with a nominalized verb (7a,b). (7a) translates word by word into German (7b), with the exception of the cliticized negation.

(7) a. Petrovi (ne)bylo do smíchu/ řeči/ zpěvu.
        Peter_{:dat} (neg)was_{:neut} PREP laugh_{:gen} /talk/sing
        'Peter did (not) feel like laughing/talking/singing'
     b. Dem Peter war (*es*) nicht nach Lachen/Reden/Singen.
        the Peter_{Dat.}was (it) not PREP laughing/talking/Singing

Such a construction is inaccessible in an [S[VO]] language since it does not contain a subject and, as German confirms, there is no (hidden) semantically empty subject involved. So, the Czech case (7b) is a case of a subjectless construction.

Let us summarize the discussion of property (ii). In the Czech sentence structure, a structural subject position is neither obligatorily present nor obligatorily filled. The contrast between Czech and undisputed SVO languages is clearly demonstrable. Czech behaves as expected & predicted for a T3 language.

Let us turn now to property (v.), viz. the absence of the LLC effect for immediately preverbal adjuncts in T3 languages. Again, Š&J contest our prediction. We predict that the LLC effect is absent in Czech because the VP does not count as a strictly head-initial VP in a T3 language such as Czech or Russian. Š&J's objection is in fact easy to dismiss. What they present is two sentence pairs, one from Czech and one from Russian. They rate one sentence of each pair as ungrammatical. In each case it is a sentence that consists of a noun and a clitic at the beginning, and a verb at the end. The 'meat' of this syntactic 'hamburger' is a *single, big, center-embedded* phrase, which could be easily extraposed. We are not surprised that such utterances would be rated less 'palatable' or even unacceptable, in comparison with the extraposed version. But, we are surprised that this is presented as a counter-argument. Š&J have only shown that utterances are avoided that contain 'very heavy' center-embedded phrases that could be extraposed. But this is not the point. The point is that, *independent of their size*, pre-VP adjuncts must be head-

adjacent to the VP in languages with strictly head-initial VPs. This is easy to test. The test is not restricted to APs. It is sufficient to adduce head-initial PPs.

A search in three big English corpora[8] confirms the effect of LLC in English. The expression "*should more* carefully" is well attested in each corpus. However, as expected for a PP in the pre-VP position, the sequences "*should with care*", "*should with great care*", or "*should with more care*" are absent in these three corpora in the pre-VP position (see Haider 2018), just like other adjunct PPs (19c). They are frequent in clause final or clause-initial positions, however.

(8) a. She (has) *much more carefully* examined the case
     b.*She (has) *with great care* examined the case
     c.*She (has) *after a few minutes* stopped the examination

It is easy to locate Czech corpus data showing that adverbial PPs occur in immediately preverbal positions (9). Note that the BNC does not contain a single token of "*before midnight*", "*with great care*" or "*with pleasure*" immediately before the verb although more than thousand tokens of each PP can be retrieved from this corpus.

(9) a. Říkala jsi, že tě Kristvin [*před půlnocí*] opustil, asi kolem jedenácté.[9]
         said you , that you$_{Obj.}$ Kristvin [*before midnight*] left, about eleven o'clock.
     b. moderní psychologie [...] *s velkou pečlivostí* definuje stovky nových termínů […].[10]
         modern psychology *with great care* defines hundreds of new terms
     c. Rudla *s radostí* slyšel o jeho úspěších […].[11]
         Rudla *with pleasure* heard about his successes.

In sum, Czech is evidently not subject to the LLC constraint. This is what is predicted if the Czech VP is directionally unconstrained, which is a defining T3 property.

Eventually, Š&J object to line (vi.) in the table, that is left-branch extractions from noun phrases in preverbal positions. The term 'preverbal' refers to the relevant positions in the clause structure of SVO language. The subject and any item fronted out of the verb phrase are 'preverbal' w.r.t. the base position of the verb in the clause. Sub-extraction out of an NP or PP is licit only out of phrases in 'postverbal' positions, that is, VP-internal positions. In SVO, the 'preverbal' ones are either spec-positions or left-adjoined positions, and the CED constraint blocks extraction out of these position; see Haegeman et als. (2014) for details.

We understand, as Š&J explain convincingly, that Czech restricts focused items to postverbal positions. In left-branch extractions, the item left behind is typically focused. Therefore, such a phrase cannot be placed in a preverbal position in Czech. This shows that an intervening independent factor prevents the testing of these left-branch extractions in preverbal positions in Czech. Nevertheless, Š&J's example (12), repeated her under (10), is telling and instructive:

(10)   Která$_1$ to upekla [t$_1$ kuchařka]?
            Which it baked cook

---

[8] BNC = British National Corpus (100 million: British, 1980s-1993); CocA = Corpus of contemporary American English (520 million: US, 1990-2015); NOW = News on the web (5.2 milliard: Web news, since 2010).

[9] Link: https://books.google.at/books?id=Ohq2DwAAQBAJ&pg=PT232&lpg=PT232&dq=%22p%C5%99ed+p%C5%AFlnoc%C3%AD+opustil%22&source=bl&ots=TkdxmP8_2p&sig=ACfU3U0cxpObMu-pulrQvXKpbLhf5na5FvQ&hl=de&sa=X&ved=2ahUKEwj2q4K9tbH0AhVoh_0HHVKtBEoQ6AF6BAgQEAM#v=onepage&q=%22p%C5%99ed%20p%C5%AFlnoc%C3%AD%20opustil%22&f=false

[10] Link: https://www.google.de/books/edition/Malign%C3%AD_onemocn%C4%9Bn%C3%AD_psychika_a_stres/qPYOE-AAAQBAJ?hl=de&gbpv=1&dq=s+velkou+pe%C4%8Dlivost%C3%ADzkou%C5%A1el&pg=PA134&printsec=frontcover

[11] Link: https://www.google.de/books/edition/Vzorek_bez_ceny_a_pan_Biskup_aneb_za%C4%8D%C3%A1/54pEAAAAMAAJ?hl=de&gbpv=1&bsq=%22s+radost%C3%AD+sly%C5%A1el%22&dq=%22s+ra-dost%C3%AD+sly%C5%A1el%22&printsec=frontcover

'Which cook baked it?'

In (10a), 'cook' is the subject of a transitive clause, and so its structural position is, according to Jasinskaja & Šimík (forthc. ex. 21b), the spec position of vP. From there, it is supposed to move to Spec T. The object in (10) is a fronted pronoun. In the paper, they do not disclose how transitive subjects would end up in a postverbal position. There are two possibilities. Either the verb (and the object) are moved to a higher position or the subject is extraposed. In the latter case, the subject ends up in a structurally high adjunct position and is thereby opaque for extraction. In the former case, the subject is opaque by virtue of being still contained in a preverbal Spec position. So (10) should be ungrammatical and unacceptable under any of these analyses, which it isn't. In the T3 analysis, the subject in (10) is in its base position *and* in the licensing domain of the verb. The verb is instantiated in one of the higher positions in the T3 VP-structure. Therefore, transparency for sub-extraction is expected.

Eventually, the programmatic title of Š&J's – "*There is no single Slavic word order type*" – invites an aside. It is in remarkable contrast to the title message of another paper of theirs, viz. "*Slavonic free word order*" (Jasinskaja & Šimík forthc.), referring to a pan-Slavonic word order property. Slavonic free word order is a word order type[12], namely T3, and this type covers all Slavic languages (except Sorbian). The Slavic word order type characterizes a type of languages with a clause structure with mobile verbal heads without directional licensing restrictions, and the concomitant potential for scrambling, in other words: It is the T3 type.

In sum, we are content to find ourselves in a position to conclude on solid empirical grounds that the evidence for classifying Czech as a Type 3 languages is good, and as good as the evidence for any other Slavic language.

Bibliography

Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed.), *Clause structure, language typology and syntactic description*, Vol. 1, 61-131. Cambridge: Cambridge University Press.

Franks, Steven 2005. Slavic languages. In Cinque, Guglielmo & Kayne, Richard (eds.) The Oxford Handbook of Comparative Syntax. 373-419. Oxford University Press.

Guiraud-Weber, Marguerite & Kor Chahine, Irina 2013. Impersonals and beyond in Slavic. In: Irina Kor Chahine (ed.) *Current studies in Slavic linguistics*. 1-21. Amsterdam: John Benjamins.

Haider, Hubert. 2018. The Left-Left Constraint – a structural constraint on adjuncts. Ulrike Freywald & Horst Simon (eds.) *Headedness and/or Grammatical Anarchy*? Berlin: Language Science Press. (submitted 2018). [lingbuzz/004970]

Haider, Hubert. 2019. On absent, expletive, and non-referential subjects. In Herbeck, Peter, Bernhard Pöll & Anne C. Wolfsgruber (eds.) *Semantic and syntactic aspects of impersonality*. 11-46. Hamburg: Helmut Buske Verlag.

Haider, Hubert. 2020. A null theory of Scrambling. (Special issue of) *Zeitschrift für Sprachwissenschaft* 39(3): 375-405. Volker Struckmeier & Andreas Pankau (ed.) "Indeterminacies and mismatches in grammatical systems." doi: https://doi.org/10.1515/zfs-2020-2019.

Haegeman, Liliane, Jiménez-Fernández, Ángel L. & Andrew Radford. 2014. Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1): 73-150. DOI 10.1515/tlr-2013-0022

Jasinskaja, Katja & Šimík, Radek (forthc.). Slavonic free word order. (forthcoming) in Fellerer Jan & Neil Bermel. *The Oxford guide to Slavonic languages*. Oxford University Press. lingbuzz/004919

---

[12] According to Dryer (2007:113), *"languages with highly flexible word order are themselves a linguistic type."*

Kuboň, Vladislav, Lopatková, Markéta & Hercig, Tomáš. 2016. Searching for a measure of word order freedom. *ITAT 2016 Proceedings, CEUR Workshop Proceedings* Vol. 1649: 11-17. (http://ceur-ws.org/Vol-1649/).

Scholze, Lenka. 2015. Deutscher Einfluss auf die obersorbische Wortfolge – Vergleich von Schrift- und Umgangssprache. In S. Wölke & H. Bartels (eds.), *Einflüsse des Deutschen auf die grammatische Struktur slawischer Sprachen*, 198-213. Bautzen/Budyšin: Domowina.

Siewierska, Anna / Uhliřová, Ludmila 2010. An overview of word order in Slavic languages. In A. Siewierska (ed.), *Constituent order in the languages of Europe*, 105-150. Berlin: Mouton De Gruyter.

Szucsich, Luka. 2006. Nothing wrong with finite T: Non-agreeing accusative impersonal Sentences. In Golędzinowska Magda, Ulyana Savchenko & Richard Compton (eds.) *Formal Approaches to Slavic Linguistics 15, The Toronto Meeting 2006*. 417-435. Ann Arbor: Michigan Slavic Publications.