1. Introduction and proposal

In this paper, we address the syntax and semantics of clausal arguments like (1) and adjuncts like (2) in two different syntactic positions: (a) left-joined and (b) integrated. All data are from Czech, unless indicated otherwise.

(1) a. Co(koliv) mu dáš, utratí.
    what(ever) him give.2SG spends
    ‘What(ever) you give him, he’ll spend (it).’

b. Utratí, co(koliv) mu dáš.
    spends what(ever) him give.2SG
    ‘He’ll spend what(ever) you give him.’

(2) a. Když odejdeš, budu smutný.
    when/if leave.2SG will.be.1SG sad
    ‘When/If you leave, I’ll be sad.’

b. Budu smutný, když odejdeš.
    will.be.1SG sad when/if leave.2SG
    ‘I’ll be sad when/if you leave.’

We put forth novel evidence from locality, suggesting that conditionals/(cor)relatives in the left-joined position are clausal (CP), denoting (sets of) propositions, and the corresponding integrated structures are nominal (NP), possibly embedded in a PP, denoting entities of the appropriate sort (individuals, situations, time intervals, etc.). More particularly, we propose that the underlying LFs of sentences like in (1)/(2) – independently of the (a)/(b) distinction – are as in (3)/(4).\footnote{The shared LF implies that (a) and (b) are truth-conditionally equivalent, which seems correct to us.}
The wh-clauses are represented twice: once as conditional antecedents (arguments of a left-peripheral “conditional-generating” operator, assuming Kratzer’s 1979 semantics for conditionals) and once as definite descriptions, donkey-anaphoric to the entity introduced in the conditional by the wh-word. The way (a) and (b) differ is in their overt form: in the left-adjoined variant, the wh-clause spells out the conditional antecedent; in the integrated variant, the wh-clause spells out the definite description, essentially being a free relative, which are, rather uncontroversially, definite (ever since Jacobson 1995; see also Šimík to appear). Our analysis is, in a sense, a generalization of Hirsch’s (2016) proposal for ever free relatives, exemplified by (1b) with ‘ever’. Hirsch proposes that ever free relatives are unconditionals (in the sense of Rawlins 2013) + donkey-anaphoric free relatives (at LF). What we propose here is that all wh-clauses are conditionals (incl. correlatives; Brasoveanu 2008; Arsenijević 2009) + donkey-anaphoric free relatives. A fuller exposition of this general claim will have to await another occasion.

In the present paper, our aim is more modest: We will show that left-adjoined wh-clauses are consistently transparent for A'-extraction, while their string-identical integrated versions are syntactic islands. This follows if the former are CPs and the latter are NPs. We propose a minimalist analysis employing phases and antilocality, in which the presence of the NP phase plays a crucial role in blocking extraction.

For reasons of space, we concentrate on syntax, leaving compositional semantics for another occasion (but see Hirsch 2016 and Šimík 2018 for analyses compatible with the present syntax). What we also leave open is the question how the LFs in (3) and (4) are arrived at, or more particularly, whether the wh-clause is represented in both positions or replaced by an unstructured variable in one of them and, if the former is the case, whether the relation is one of movement (or multidominance) or base-generated “matching” followed by ellipsis. Finally, let us point out that while all our data are from Czech, an initial informal survey suggests that the pattern extends to other Slavic languages (not reported here).

The rest of the paper is structured as follows. Section 2 is a systematic argument showing that left-adjoined wh-clauses are transparent, while integrated ones are not. Section 3 proposes an analysis of the observed extraction pattern. Section 4 concludes the paper.

2. Evidence from locality

We will show that the pattern schematized in (5), where XP has undergone A'-movement out of the wh-clause, abbreviated as C, holds in Czech. The pattern supports our claim

---

2 Hirsch proposal concerns English ever free relatives expressing ignorance. Šimík (2018) extends Hirsch’s analysis to other languages and other semantic types of ever free relatives.
that C = CP in (5a) and C = NP in (5b). Further independent evidence will be used to support the assumption that NP/DP-hood is the factor behind the island-status of wh-clauses.

(5)  
   a.  [XP₁ [C … t₁ …] main clause]  
   b.  * [XP₁ main clause [C … t₁ …]]

Consider examples (6)–(8), all of which exemplify the pattern in (5). The extracted element is a relative pronoun, which takes part in the derivation of a headed relative clause. (The intended interpretations of the ungrammatical (b)-examples are the same as of the corresponding (a)-examples.)

(6) Extraction from (a) correlative vs. (b) free relative
   a. To je ten chlap, [ kterému₁ [C co(koliv) dáš t₁], to ztratí].
      it is the man.NOM which.DAT what(ever) give.2SG that lose
      ‘This is the man such that he will lose whatever you give him.’
   b.* To je ten chlap, [ kterému₁ ztratí [C co(koliv) dáš t₁]].
      it is the man.NOM which.DAT lose what(ever) give.2SG

(7) Extraction from (a) left-adjoined vs. (b) integrated conditional
   a. Potkal člověka, [ kterého₁ [C když poprosíš t₁], tak to zařídí].
      met man.ACC which.ACC when ask.2SG so it arrange
      ‘He met a man such that he/she/they will arrange it if you ask him.’
   b.* Potkal člověka, [ kterého₁ to zařídí [C když poprosíš t₁]].
      met man.ACC which.ACC it arrange when ask.2SG

(8) Extraction from (a) left-adjoined vs. (b) integrated purpose clause
   a. To je řečník, [ kterého₁ [C aby nalákal t₁], museli by mit peníze].
      it is speaker which.ACC so.that attract must sbj have money
      ‘This is a speaker such that they need more money in order to attract him.’
   b.* To je řečník, [ kterého₁ by museli mit peníze [C aby nalákal t₁]].
      it is speaker which.ACC sbj must have money so.that attract

We proceed as follows. We demonstrate (i) that the extraction site is within C rather than in the main clause, (ii) that the landing site is outside of C (not just at its edge) and, more particularly, in the extended projection of the main clause, and finally (iii) that what blocks the extraction in cases like (b) above is the presence of a nominal layer on top of C.

---

3 The pattern is surprising because it goes in the opposite direction to what many influential approaches to locality would predict, including theories based on Condition on Extraction Domains (Huang 1982, Chomsky 1986, Manzini 1992), which prohibit extraction from adjuncts and/or derived positions, or theories based on information structure (Erteschik-Shir 1973, Erteschik-Shir & Lappin 1979), which prohibit extraction from “non-dominant” (topicalized, presuppositional) constituents.

4 This pattern was first discussed by Lešnerová & Oliva (2003), who provide many corpus examples, as well as an analysis within a dependency approach to syntax.
2.1. Extraction site is in C

Let us consider example (6) and assume, for the sake of the argument, that the relative pronoun has been extracted from the main clause subject position. The subject is indeed covert and coreferential with the relative pronoun, so it could well be a trace, while the gap internal to C (the indirect object of ‘give’) could be parasitic. Such an analysis would indeed conform to much of what we know or believe about how natural language syntax works. Yet, it is demonstrably wrong. First, the relative pronoun carries dative – assigned within C – and not nominative, as expected under our strawman analysis. A nominative relative pronoun is in fact ungrammatical; see (9a). Second, if the main clause gap corresponds to an oblique-case-marked argument, it is ungrammatical, as shown in (9b), where the argument position must be filled by an overt (clitic) dative pronoun. Third, the main clause need not contain an occurrence of a gap or pronoun coreferential with the extracted element. This is evident from (8) and reiterated in (9c). We conclude that if there is a subject gap coreferential with the extracted element, as in (6a)/(9a) or (7), then it is a plain vanilla pro, not a trace.

(9) a. To je ten chlap, { kterému / * který} [C co(koliv) dáš ], to ztratí. ‘This is the man which.give.2SG that lose.2SG’
   ‘This is the man such that he will lose whatever you give him.’

   b. To je ten chlap, kterému [C když zavoláš ], pomůžeš { mu / *Ø}. ‘This is the man such that if you call him, you’ll help him.’

   c. To je ten chlap, kterému [C když zavoláš ], ( tak) odejdu. ‘This is the man such that if you call him, I’ll leave.’

Example (10) shows that the extracted element reconstructs into C for purposes of reflexive anaphor binding (svou ‘his.REFL’ is bound by Karel), which is known to be subject to stringent locality conditions in Czech (Dočekal 2005). Notice that while the main clause has a pro coreferential with Karel, there is no position for the extracted phrase to reconstruct to and hence no way that svou ‘his.REFL’ could be bound by this pro. (The relevance of the two interpretations will become evident shortly.)

(10) [ Kterou svou, cennost] říkal, že [C když si Karel uschová t1 ], which his.REFL valuable said.PL that when REFL Karel deposits
tak udělá nejlíp? so does best
   a. Available interpretation: ‘Which one of Karel’s valuables is such that they said that it would be wise of Karel if he deposits it?’

5 Main clause gaps are sometimes licensed (and attested in corpus) if they correspond to accusative-marked pronouns (Lešnerová & Oliva 2003: 249). This is arguably related to the fact that accusative-marked resumptive pronouns sometimes alternate with gaps in Czech resumptive relatives (see e.g. Toman 1998).
Structure of conditional and (cor)relative clauses

b. *Unavailable interpretation:* ‘Which one of Karel’s valuables is such that it would be wise of Karel if they said that he deposits it.’

2.2. **Extraction targets a position external to C**

Believers in the ban on extraction out of (left) adjuncts might have thought – at least until encountering our last example – that the extracted element has not left the adjunct in the first place – perhaps it just moved to the edge of the adjunct. This assumption has in fact been adopted by Heck (2008) or Grewendorf (2015), in order to make sense of comparable constructions in (Bavarian) German, exemplified in (11).

(11) De Mass wenn i no drink, bin i bsuffä. (Grewendorf 2015: 232–233)
the liter if I still drink am I drunk
‘If I still drink this Mass, I’ll be drunk.’

Example (10) clearly shows that the extracted phrase can escape C in Czech. Importantly, as the unavailability of interpretation (10b) explicates, the verb říkali ‘said’ embeds the main clause (modified by C), not just C. Similar kind of evidence is provided by the placement of epistemic and evidential particles, such as prý ‘allegedly’, exemplified in (12). As the availability of the interpretations in (a)/(b) suggests, prý is attached to (semantically modifies) the main clause (MC), not the conditional. In fact, in Czech, as in English, epistemic and evidential adverbs cannot even modify conditional antecedents (for English, see McDowell 1987, Drubig 2001), which is also the reason why we have replaced *allegedly* with they say that in (12b). We conclude that the landing site of elements extracted out of C is, or at least can be, external to C.

(12) To je ten člověk, který MC prý MC [C když promluví], všichni ztichnou.

a. *Available interpretation:* ‘This is the man who allegedly speaks everybody falls silent when he speaks.’

b. *Unavailable interpretation:* ‘This is the man such that everybody falls silent if they say that he speaks.’

2.3. **What blocks extraction is a nominal layer**

In this section, we show that it is the nominal element that blocks extraction out of C. Compare the grammatical (12) with the ungrammatical (13), in which C is embedded in the v ‘in’ PP.

(13) *To je ten člověk, kterého:1] prý [PP v ten moment [C když uvidíš t1]],

it is the man who:ACC allegedly in the moment when see.2SG will.2SG be.amazed
Example (14) shows that ungrammaticality also arises when the dominating NP is not embedded in a PP. Compare the grammatical extraction in (6a) with the ungrammatical one in (14), with the pronoun *to* ‘it’.

(14) *To je ten chlap, [ kterému to [NP co dáš t], to ztratí].
  it is the man which. it what give.2SG that lose.3SG

The following data show that the nominal layer blocks extraction irrespective of whether extraction happens from a left-adjoined wh-clause, as in (13) and (14), or from an integrated wh-clause, as in (15) and (16).

(15) *To je ten člověk, kterého [prý budeš žasnout [PP v ten
  it is the man who. allegedly will.2SG be.amazed in the
  moment [C když uvidíš t]].
  moment when see.2SG

(16) *To je ten chlap, [ kterému ztratí [NP to [C co dáš t]].
  it is the man which. lose.3SG it what give.2SG

The proposal is also supported by the fact that adjuncts introduced by subordinators containing a nominal element like *protože* ‘because’ (lit. for.it.DECL-COMP) *poté co* ‘after’ (lit. after.it REL-COMP, cf. Biskup 2019) also do not allow extraction, as demonstrated below.

(17) *To je ta knížka, [ kterou [ { protože / poté co} koupil t ]] už
  it is the book which.ACC because after bought already
  NEG.have.3SG money

3. A syntactic analysis of the extraction pattern

In this section, we provide a particular syntactic analysis of the extraction pattern observed above, building on the idea that left-peripheral clausal adjuncts are CPs, but their integrated variants are NPs, proposed in section 1 and empirically supported in section 2.

3.1 Extraction from the integrated/NP-headed wh-clause

Following Abels (2003), Grohmann (2003), Ticio (2003) and Bošković (2005), we assume that phrasal movement observes antilocality, specifically, that phrasal movement must cross at least one full phrase. In addition, we assume that Czech, as an article-less language, has the following phases: CP, vP and NP (cf. Chomsky 2007, 2008; Dyakonova 2009; Bošković 2014). In earlier generative frameworks, the impossibility of extraction from a wh-clause embedded in an NP would be explained in terms of Subjacency (Chomsky 1973) or Barrierhood (Chomsky 1986), with CP and NP being bounding nodes/barriers. The phase model can mimic it with two adjacent phases (CP
and NP), where movement from the lower phase edge to the higher phase edge violates antilocality (see Bošković 2017). This means that the illicit movement from the integrated wh-clause looks like (18).

\[ \ldots \text{[NP whP}_1\text{[NP (demonstrative) [NP N [CP t}_1\text{ C } \ldots \text{t}_1\ldots]]]} \]

The moving element (whP \(_1\) in (18)) must move out of the integrated clause via the CP-edge. Since the next higher phase edge is the edge of NP and there is no full phrase on the way (which can be crossed), the CP-to-NP movement violates antilocality. How this proposal applies to the data under discussion is shown below. As an illustration consider the left-peripheral NP-embedded conditional in (19), repeated from (13), and the integrated conditional in (20), repeated from (15). In both cases, the problematic traces (the illicit movement step) are in italics. As we proposed in section 1, the integrated case – here (20) – is always represented as an NP (and interpreted as a donkey-anaphoric entity), even if an overt nominal (ten moment “the moment”) is missing.

(19) *To je ten člověk, kterého prý v ten moment [CP \text{t}_1 \text{když uvidíš \text{t}_1}], budeš žasnout.*

(20) *To je ten člověk, kterého prý budeš žasnout v ten moment [CP \text{t}_1 \text{když uvidíš \text{t}_1}],*
The whP *kterého* moves to the edge of CP₁ and the adjunct merges with the conditional operator and functions as its restrictor.⁶ In the next step in (23), CP₁ adjoins to the matrix TP (the nucleus). After that, *kterého* is going to move to CP of the matrix clause. As discussed in the preceding section, this step must cross at least one full phrase in order not to violate antilocality. Given that phrasal elements like the adverbial *prý* can intervene between the moved *který* and the adjunct clause, this seems to be the case (we assume a Mood projection for such adverbials). And since the wh-clause in (22) is not embedded in an NP, the movement step is licit.

(23)    NP    3
clověk  NP     CP    3
‘man’  ktere
whP₂   ‘who’
      whP₂
      ‘who’

4. Conclusion

We showed that A’-extraction in Czech is possible out of left-peripheral clausal adjuncts (conditionals, correlative, and the like), but impossible out of string-identical, but syntactically integrated clauses. We argued that this surprising extraction pattern – previously observed in Lešnerová & Oliva 2003, but novel to the generative discourse – follows from the idea that left-peripheral clausal adjuncts are CPs (semantically propositions), while their integrated versions are NPs (semantically entities). We

⁶ For the sake of explicitness, we assume that in the optional NP *moment*, the adjunct clause is represented as a clausal *pro*. 
proposed an analysis couched in a phase- and antilocality-based approach to syntactic islands.

References


Bošković, Željko. 2014. Now I’m a phase, now I’m not a phase: On the variability of phases with extraction and ellipsis. Linguistic Inquiry 45(1): 27–89.


