Revisiting the Person Case Constraint in Czech

Anne Sturgeon, Harvard University
Boris Harizanov, UC Santa Cruz
Maria Polinsky, Harvard University
Ekaterina Kravtchenko, UC Santa Cruz
Carlos Gómez Gallo, Harvard University
Lucie Medová, University of Southern Bohemia
Václav Koula, Charles University

1. Background and motivation

The Person Case Constraint (PCC) is a co-occurrence restriction on certain combinations of phonologically weak arguments of ditransitive verbs which is widely attested cross-linguistically (Perlmutter 1971, Bonet 1991, Anagnostopoulou 2003, Nevins 2007, a.o.). The following varieties of the PCC have been recognized in the literature:

(1) The Person Case Constraint
   In a combination of a direct object and an indirect object:
   a. Strong: the direct object has to be 3rd person
   b. Weak: if there is a 3rd person, it has to be the direct object
   c. Me-First: if there is a 1st person, it has to be the indirect object
   d. Strictly Descending: the argument with the higher person specification (where 1 is higher than 2 is higher than 3) has to be the indirect object

The strong version of the PCC prohibits combinations of 1st and 2nd person (local) objects while the weak version allows such combinations, when the indirect object is also a local person. An additional implication

* We would like to thank Judith Aissen, Barbara Citko, Hana Filip, Adam Milton Morgan, and the audiences of the UC Berkeley Syntax Circle, FASL 19, FDSL 8.5 and the 2011 LSA Annual Meeting for their constructive comments and suggestions. Also, thank you to the anonymous FASL reviewers for their constructive insights. This research was supported in part by funding from the National Heritage Language Resource Center at UCLA and from the Faculty of Arts and Sciences at Harvard University and the Davis Center at Harvard. All errors are our sole responsibility.
associated with the Me-First condition is that, in combinations of local persons, a 2nd person pronoun must be the direct object (see also Fassi-Fehri 1988; Perlmutter 1971; Fernández-Soriano 1999; and Nevins 2007, 2008). The strictly descending PCC is the version of the PCC exhibited in Czech; descriptively, it is a combination of the weak PCC and the Me-First condition.

The PCC targets phonologically weak elements only. Thus, it applies to agreement markers in languages that have three-way agreement systems where the verb agrees with the subject, direct object, and indirect object. It also applies to pronominal clitics, as in the Romance languages, and to weak pronouns in English (Perlmutter 1971, Bonet 1991, Anagnostopoulou 2003, a. o.). It is not, however, a semantic restriction. Languages with the PCC use full pronominals or prepositional phrases to express semantic meanings that cannot be expressed by weak elements.

There has been disagreement in the literature about whether or not Czech exhibits PCC effects (YES: Vos and Veselovská 1999, Franks and King 2000, Rezac 2005, Bhatt and Simik 2009, Medová 2009; NO: Lenertová 2001, Haspelmath 2004, Migdalski 2006, Hana 2007). We conducted two experimental studies to probe this issue in Czech: an acceptability-rating experiment and a preliminary corpus study. Both studies showed that Czech exhibits what descriptively appears to be the Strictly Descending PCC.

Example (2a) illustrates the weak version of the PCC in Czech. A 1st person indirect object clitic co-occurs with a direct object 3rd person clitic, however a 1st person clitic cannot be a direct object in combination with a 3rd person indirect object.

(2) a. Karel mi ji /mu mě ukázal
   Karel 1SG.DAT.CL 3SG.ACC.CL 3SG.DAT.CL 1SG.ACC.CL showed
   na fotce.
   on photo
   ‘Karel showed her to me in the photo.’

Abbreviations are as follows: AUX (auxiliary), CL (clitic), SG (singular), PL (plural), ACC (accusative), DAT (dative).

1 In the notation <x y>, x and y are the person specification of the indirect and direct object respectively.
Czech also exhibits the Me-First condition, (2b). In combinations of local persons, the first person must be the indirect object.

(2) b. Představil mí tě /*tí mě
introduced 1SG.DAT.CL 2SG.ACC.CL/ 2SG.DAT.CL 1SG.ACC.CL
včera v Hradci Králové.
yesterday in Hradec Králové
‘He introduced you to me yesterday in Hradec Králové.’
√1<2>, *(<2 1>

The paper is organized as follows. In section 2 we present background on the Czech clitic system. Our experimental and corpus research is presented in sections 3 and 4. We explore Anagnostopoulou 2003 and Nevins’s 2007 Multiple Agree approach to the PCC in section 6. The discovery that the non-canonical Acc > Dat clitic order emerges as grammatical only in otherwise PCC-violating clitic clusters (e.g. 1Acc > 3Dat) presents a certain difficulty for such approaches. In section 7 we suggest the possibility that Czech, in fact, is not subject to the PCC. Instead, we demonstrate that the observed (un)grammaticality patterns could be explained in terms of linearization constraints on the clitic cluster.

2. Background on Czech clitics

Czech has both pronominal and verbal clitics which cluster together in the “second position” within a clause following the order Dat > Acc. It is generally assumed that second position clitics syntactically appear in a high functional head (for us, T₀) (Fried 1994, Veselovská 1995, Franks and King 2000, Lenertová 2001). The inventory of Czech pronominal clitics is given in (3); we will not consider reflexives here.

(3) Czech pronominal clitics

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>3SG-M</th>
<th>3SG-F</th>
<th>1PL</th>
<th>2PL</th>
<th>3PL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acc</strong></td>
<td>mě</td>
<td>tě</td>
<td>ho</td>
<td>ji</td>
<td>nás</td>
<td>vás</td>
<td>je</td>
</tr>
<tr>
<td><strong>Dative</strong></td>
<td>mí</td>
<td>ti</td>
<td>mu</td>
<td>jí</td>
<td>nám</td>
<td>vám</td>
<td>jim</td>
</tr>
</tbody>
</table>
Czech also has a series of strong pronominal elements, shown in (4). Many of the strong pronouns are identical orthographically to the clitic forms.

(4) Czech strong pronouns

<table>
<thead>
<tr>
<th></th>
<th>1SG</th>
<th>2SG</th>
<th>3SG-M</th>
<th>3SG-F</th>
<th>1PL</th>
<th>2PL</th>
<th>3PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accusative</td>
<td>mě</td>
<td>tebe</td>
<td>jeho</td>
<td>jí</td>
<td>nás</td>
<td>vás</td>
<td>je</td>
</tr>
<tr>
<td>Dative</td>
<td>mne</td>
<td>tobě</td>
<td>jemu</td>
<td>jí</td>
<td>nám</td>
<td>vám</td>
<td>jim</td>
</tr>
</tbody>
</table>

Full pronouns, as opposed to clitics, appear in contrastive topic and focus positions at the left and right edges of the clause, respectively, and are associated with specific prosodic contours (Hajičová et al. 1995).

3. Acceptability rating study

3.1 Methodology

In an online acceptability rating study, 143 native speakers of Czech rated sentences they read on a scale from 1 (acceptable) to 7 (unacceptable). There were 6 conditions and 16 experimental sets (96 experimental items and 276 fillers). The following ditransitive verbs were used in the study: čválit ‘praise’, představit ‘introduce’, doporučit ‘recommend’, and ukázat ‘show’. All combinations of plural and singular clitics were used in the study: <1 2>, <2 1>, <1 3>, <3 1>, <2 3>, <3 2>. Sample stimuli are shown in (6):

(5) a. Subject – DAT.CL – ACC.CL – Verb – Adjunct
    b. Verb – DAT.CL – ACC.CL – Adjunct – Adjunct

(6) a. Vedoucí mi ho doporučil minulý týden.
    chief 3PL.DAT.CL 1PL.ACC.CL recommended last week
    ‘The boss recommended us to them last week.’ <1 3>

3 We did not use <3 3> combinations because we assume that the <3 3> clitic combination is fully grammatical.
b. *Doporučil mu mě na místo minulý týden.
   ‘He recommended me to him for the position last week.’  <3 1>

3.2 Results
The results (calculated within a linear mixed model fit by restricted maximum likelihood, REML) are shown in Figure 1.

Figure 1: Acceptability rating task results; x-axis: person combinations; y-axis: the likelihood of a sentence containing a given person combination to be rated higher/lower than the baseline (<1 2>).

The main results are as follows:

(7) a. <3 1> and <3 2> violate the Weak PCC and are rated lower
b. <2 1> and <3 1> violate the Me-First PCC and are rated lower
c. \(<1\ 3>\) and \(<2\ 3>\ do\ not\ violate\ the\ PCC\ and\ are\ rated\ higher\)

Czech thus exhibits what descriptively looks like the Strictly Descending PCC (the combination of the Weak PCC \((1b)\) and the Me-First Condition, \((1d)\)) with both plural and singular clitics. A similar result has also been reported in Classical Arabic (Fassi-Fehri 1988 and Nevins 2007, 2008), as well as in some dialects of Spanish (Perlmutter 1971, Fernández-Soriano 1999).

4. A corpus study

Using SYN2005 (part of the Czech National Corpus (CNC)) and the web (http://www.google.com and http://www.seznam.cz), we investigated the naturally occurring distribution of dative and accusative clitic pairs. The findings of the corpus study mirrored the results from the judgment study. Clitic clusters that violate the Strictly Descending PCC are rare.

4.1 The Acc > Dat order obviates PCC effects

Medová 2009 was the first to note that the otherwise ungrammatical Acc > Dat order obviates PCC violations. It is important to note that this clitic order only surfaces when the clitic combination violates the PCC. We have a total of 12 examples from the web that involve this unexpected order.

\( \begin{align*}
(7a, b): & \ \ p<.0001, \\
(7c): & \ \ p<.0002.
\end{align*} \)

We would like to attribute any gradience observed among \(1>3,\ 2>3\) and \(1>2\) not to a categorical difference in grammaticality but to extra-grammatical factors, such as frequency.

These results are not influenced by orthographic overlap between strong and clitic pronouns (see \((3-4)\)). The results are the same when only the clitic forms which are segmentally identical to strong forms are considered.

SYN2005 is part of the Czech National Corpus (CNC), which can be found on http://ucnk.ff.cuni.cz. It is a balanced corpus of written Czech of 100 million words. Most of the texts are from 2000-2004 (fiction: 40\%, technical literature: 27\%, journalism: 33\%).

Future corpus work is necessary to determine whether the low frequency of PCC violations is due to their ungrammaticality or to the overall low frequency of certain combinations of arguments (e.g. 3\textsuperscript{rd} person dative and 1\textsuperscript{st} person accusative). This will require collecting frequency counts of combinations of non-clitic arguments (cf. Haspelmath 2004).
clitic ordering, two of which are shown in (8).\footnote{Further work is necessary to determine whether this order is possible with local person combinations only. No examples were found in a preliminary corpus study.}

\[ (8) \]
\begin{enumerate}
\item a. \ldots and with pleasure I passed me to him.' \url{http://www.stmivani.eu/36-jednodilne-povidky/vlkodlaci-svatba-novy-mesic/}
\item b. \ldots I won’t give you to him!’ \url{http://ff-cole-dylan.blog.cz}
\end{enumerate}

Elicitation of primary data reveals that speakers tend to strongly prefer this order of clitic pronouns (Acc > Dat) when the PCC is violated.

4.2 Attested PCC violations
Languages with the Weak PCC show considerable inter- and intra-speaker variation (see Bonet 1991, Anagnostopoulou 2005, Nevins 2007, Ormazabal and Romero 2007). We also found this to be the case for Czech. A total of 10 examples in which the PCC is violated were found (2 from the CNC and 8 from the web), see (9).

\[ (9) \]
\begin{enumerate}
\item a. \ldots I’ll give you to him in a bag.
\item b. \ldots bring you to him by force…
\end{enumerate}

Sentences which contain PCC-violating clitic combinations, though attested in corpora, are often judged as ungrammatical or degraded by native speakers.
5. A syntactic approach to the PCC

Any theoretical approach needs to account for the Strictly Descending PCC, see (1d). In this section a syntactic analysis of this generalization is proposed. We conclude by raising the question of the difficulty of capturing the alternative Acc > Dat orders in a strictly syntactic account. Section 6 suggests a different view: Czech is not subject to the PCC and what looks like the Strictly Descending PCC on the surface, in fact, results from constraints on the linearization of clitic clusters.

5.1 Assumptions and background

Most syntactic accounts suggest that the PCC arises in contexts in which there are “two arguments against one head” (Anagnostopoulou 2003, Béjar and Rezac 2003, Adger and Harbour 2007). Anagnostopoulou’s proposal is that strong and weak PCC effects arise when the dative and accusative arguments of a transitive verb enter into a feature-checking relationship with the same functional head, in this case, T₀. To account for the Weak PCC, she argues that Multiple Agree in double argument configurations is established between the probe, T₀, and the two goals (Ura 1996, Hiraiwa 2004), and that the person features are checked simultaneously against both the indirect and direct objects. Ungrammaticality arises when there is a clash between the feature specifications of the indirect and direct objects.

Anagnostopoulou follows Taraldsen 1995 and Ritter 1995, a.o., in assuming that third person pronouns are ‘determiner pronouns.’ Unlike them, however, she argues that some third person pronouns are specified for person: indirect object 3rd person is [-person] (direct object 3rd persons have no person specification).

The grammaticality of <1 3> and <2 3>, then, arises from the fact that there is no clash in person specifications between the 2 objects; the indirect object is marked for person ([+person]) and the direct object has no person specification. The direct object can check number without interference from the indirect object because, Anagnostopoulou assumes,

---

9 This is only relevant for clitic arguments; full pronouns and prepositional phrases do not need to check features with a higher functional head.
Dative arguments lack a number specification (see Taraldsen 1995 for more details).

The ungrammaticality of <3 1> and <3 2> combinations results from restrictions against conflicting feature specifications between the two objects. In particular, a [-person] feature on the third person dative argument and a [+person] feature on the first or second person accusative argument results in a feature clash of person. The Weak PCC also allows the combinations <1 2> and <2 1>, as there is no clash in person specification. We encounter a problem at this point. Anagnostopoulou’s feature system is not adequately nuanced to capture the Me-First condition; while <1 2> is a grammatical person combination, <2 1> is not. We therefore need a richer representational vocabulary to distinguish between 1st and 2nd persons.

5.2 Deriving the strictly descending PCC
The following matrix of person features makes the necessary distinctions (Nevins 2007, 2008, a.o.):

(10) Person features
    a. [+author] true iff the reference set contains the speaker;
    b. [+participant] true iff the reference set contains one of the discourse participants.

(11) Person representations

<table>
<thead>
<tr>
<th></th>
<th>author</th>
<th>participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Person</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2nd Person</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3rd Person</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

This set of features allows us to distinguish between local persons and account for the Me-First condition exhibited in Czech (as well as other languages).

Nevins 2007 argues that the PCC arises when both the clitics are in the same agreement domain: the domain of a single, probing head in this

---

10 [+author, -participant] is logically impossible because [+author] is a proper subset of [+participant].
case is $T^0$. Nevins proposes that syntactic agreement is relativized to the marked values of [author] and [participant]. Under this proposal, the probe $T^0$ is relativized to [+author] and [+participant]; those are the only values that are visible to $T^0$. In all these respects his proposal is similar to the one by Anagnostopoulou. Unlike Anagnostopoulou, Nevins assumes that Multiple Agree always applies and that the different varieties of the PCC all result from locality constraints on its application. In particular, to account for the strictly descending PCC he proposes that a convergent derivation requires that there be no unmarked values of [author] or [participant] that intervene between a probe and a goal that have marked values for author or participant. The intuition behind this proposal is that, given the features [+/-author], [+/-participant], a marked value of either of these features cannot be only on the lower argument within the domain of $T$.

Examples of convergent and non-convergent derivations are given in (12a-b). (12a) illustrates a convergent derivation for $<1 3>$. No marked values for [author] or [participant] appear on the accusative argument. The dative argument is marked for both [author] and [participant] (1st person), while the accusative argument is unmarked for both. Derivations involving $<3 3>$ and $<2 3>$ proceed similarly. In the case of $<2 3>$, marked values for [participant] only appear on the higher argument (the dative). In the case of $<3 3>$, the derivation converges because there are no marked values on any argument.

(12) a. $1 > 3$: $\sqrt{TP \, [vP \ldots DAT^{[+auth, +part]} \ldots ACC^{-auth, -part}]}$

b. $*3 > 2$: $*TP \, [vP \ldots DAT^{-auth, -part} \ldots ACC^{+auth, -part}]$  

(12b), on the other hand, demonstrates that the derivation for $<3 2>$ does not converge. Marked values appear only on the lower argument: the local 2nd person. A similar configuration would be found with $<3 1>$. Both the Weak PCC and the Me-First condition fall out from this analysis.

5.3. Syntactic approach and Acc > Dat order
There are at least two analytical paths we could take under a syntactic approach. First, it could be assumed that there are two different base-generated configurations, one where the dative argument is higher and
one where the accusative is higher. However, Dvořák (2009) provides evidence that for the majority of ditransitive verbs in Czech the dative argument is higher than the accusative one. Another option would be to suggest that movement of the accusative argument takes place before Multiple Agree is established. Whatever modification to the syntactic analysis is introduced, it must provide an explanation of why the alternative order, Acc > Dat, is only available in the case of PCC-violating clitic combinations; the syntax must be aware of the relative person specifications of the two internal arguments of the verb. At this point we would like to suggest a different way of viewing the (un)grammaticality patterns observed in Czech.

6. Czech is not subject to the PCC

What we are proposing for Czech is that what descriptively amounts to the Strictly Descending PCC is not actually the PCC, but, rather could be viewed as the interaction of the following two constraints on the linearization of complex heads (regardless of their syntactic position):

(13) **Generalizations (≠ Strictly Descending PCC (1d))**

a. The clitic argument with the “higher” person specification (where 1 is higher than 2 is higher than 3) has to precede the other argument.

b. The Dat clitic argument has to precede the Acc argument unless this interferes with the condition in (a).

We can express these two generalizations in an Optimality Theoretic (OT) framework using person and case hierarchies (Prince and Smolensky 1993, Legendre 2000).

(14) a. **Edgemost (π):** The clitic with person specification π is at the left edge of the clitic cluster.

---

11 Interestingly, Dvořák identifies a limited class of verbs for which it is the accusative argument that is higher. If the PCC effects in Czech are due to intervention effects, the prediction is that these verbs will behave differently with respect to the PCC from the verbs for which the dative is higher.
b. *Edgemost* (\(\partial\)): The clitic bearing the case \(\partial\) is at the left edge of the clitic cluster.

The Czech data can then be modeled by ranking the importance of adhering to each of the hierarchies. Observing the person hierarchy (1 > 2 > 3) is more highly ranked than observing the case hierarchy (Dat > Acc). The complex head, \(T^0\), serves as input to OT evaluation, which linearizes the clitics contained within it.\(^\text{12}\)

(15) \([\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_]\)

The tableaux in (16-17) illustrate the analysis – in particular, (17) shows the emergence of the Acc > Dat order.\(^\text{13}\)

(16) Combinations that do not violate the PCC
a.

<table>
<thead>
<tr>
<th>&lt;1.DAT 3.ACC&gt;</th>
<th>E(1)</th>
<th>E(2)</th>
<th>E(3)</th>
<th>E(DAT)</th>
<th>E(ACC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. + 1.DAT 3.ACC</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. 3.ACC 1.DAT</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. + 3.DAT 3.ACC</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. 3.ACC 3.DAT</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

(17) Combinations that violate the PCC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 3.DAT 1.ACC</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. + 1.ACC 3.DAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{12}\) We assume that syntax does not provide linearization statements.

\(^\text{13}\) Note that additional faithfulness constraints are needed to prevent operations such as deletion from being a licit PCC repair strategy.
7. Conclusion

There has been considerable empirical disagreement with respect to the PCC across languages, Slavic languages in particular. Above all, in languages with the Weak PCC, judgments tend to vary across and within speakers. To address this empirical issue, large-scale acceptability studies, such as the one shown here for Czech, are needed.

As shown by our experimental studies and corpus analysis, Czech descriptively appears to exhibit the Strictly Descending PCC. However, in light of the emergence of the otherwise ungrammatical Acc > Dat clitic order, we suggested that Czech exhibits, instead of the PCC, two clitic linearization preferences: 1 > 2 > 3 (person hierarchy) and Dat > Acc (case hierarchy). If, in Czech, the person hierarchy is ranked above the case hierarchy, the empirical data can be understood. Thus, the Acc > Dat clitic ordering emerges as the grammatical way to satisfy the linearization constraints.¹⁴

References


¹⁴ Ethical datives and other non-argumental datives are not subject to the restriction described in the paper. Under the view suggested here, it must be assumed that at the point of linearization argument and non-argument datives are distinct representationally. Thanks to an anonymous reviewer for raising this point.


