Entering the sixth decade

Anna Siewierska

Local pronouns in ditransitive scenarios: Corpus perspectives from English and Polish

Abstract: While there is no doubt that the Rs (recipients) of ditransitive clauses are much more likely to be rendered by LPs (local persons) than Ts (themes), as reflected in the Person Role Constraint (see e.g., Perlmutter 1971; Bonet 1991, 1995; Haspelmath 2004), this paper sets out to determine whether the preference for LP Rs holds only relative to Ts or is in fact an absolute preference, i.e., a general property of Rs in ditransitive clauses. The investigation is based on a consideration of corpus data from two languages, English and Polish, both of which have a rich array of ditransitive predicates and together cover a good range of ditransitive constructions differing in degrees of grammaticalization. The investigation reveals that distribution of LPs is strongly affected by the degree of grammaticalization of a ditransitive construction and the encoding of the person forms, the levels of LP Rs and the frequency effects of the Person Role Constraint being strongest in double object constructions (DOC), less strong in the dative-accusative construction (DAC) and weakest in the prepositional construction (Prep-Cs), as shown in the cline: DOC > DAC > Prep-C. Significant differences are also noted in the distribution of LPs between different classes of ditransitive predicates, the highest level of LP Rs being with communicated message predicates and the lowest with ballistic and accompanying motion predicates.

Keywords: ditransitive constructions, pronouns, person hierarchy, contrastive corpus analysis

1 Introduction

As documented in Margetts and Austin (2007) languages display a wealth of strategies for encoding three participant events, among which the ditransitive construction with three explicit semantic arguments, and agent (A), theme (T) and recipient (R) may be just one of several possibilities. In European languages, however, ditransitive constructions are arguably the dominant means of encoding
three participant events and as such have received an enormous amount of attention, especially in recent years. An issue that has engendered special interest is the person restrictions on the R-T combinations that ditransitive constructions exhibit specifically in regard to the distribution of first and second persons i.e., local persons (LPs).

In what is taken to be a prototypical ditransitive construction (see e.g., Blansitt 1973; Sedlak 1975; Goldberg 2005; Malchukov et al. 2010; Bickel et al. 2010), the T is inanimate and the R human. The inanimacy of the prototypical T makes it highly unlikely to be rendered by LPs, which are by definition human. Rs, by contrast, can quite happily accommodate LPs. The association between Ts and inanimacy is not absolute, as many languages have some ditransitive predicates that allow human Ts. Even the most prototypical of ditransitive predicates, give, may do so, as is the case in English though mainly in connection with child birth (e.g., She gave him a beautiful baby boy) or matrimony (e.g., They gave him their daughter in marriage.). It has been observed that some languages impose person restrictions on the possible T and R combinations with human Ts, be it only when both the T and R are manifested as reduced pronouns, i.e., affixes or clitics as compared to independent personal pronouns. The person restrictions center on LP Ts. In some of the languages in question (e.g., French, Greek or Basque) LP Ts are precluded outright irrespective of the person of the R, as shown in (1a), in others (e.g., Catalan and Italian) only when the R itself is not an LP, as in (1b), and in yet others (e.g., Czech) only if the R is not a first person, as in (1c).

(1) Precluded Permitted
a. *T = 1/2, R = 123 a'. T = 3, R = 123
b. *T = 1/2, R = 3 b'. T = 3, R = 123; T = 1/2, R = 1/2;
c. *T = 1/2, R = 23 c'. T = 1/2, R = 1

The above restrictions on reduced pronounal LP Ts when occurring in combinations with reduced pronominal Rs, the identification of which is typically attributed to Perlmutter (1971) or even earlier to Jespersen (1927:287), are commonly referred to in the literature as the Person Role Constraint (PRC).1

Although the PRC is seen to apply to reduced pronouns but not full personal pronouns and significantly not in all languages, in a recent paper Haspelmath (2004) argues that it can actually be observed universally for all types of

---

1 The PRC is discussed also in, among others: Farkas and Kazazis (1980:77–79); Bonet (1991, 1995); Cardinaletti (2007); Nevins (2007); Ormazabal and Romero (2007); Simonenko (2008); Săvescu-Ciucivara (2009); Sturgeon et al. (2010). The formulation of the constraint varies and not all scholars distinguish between (1b) and (1c); for example Haspelmath (2004) does not.
person forms though as a preference as opposed to an absolute constraint. In other words, even in languages which do not have reduced pronominals or those which do have them but do not preclude reduced LP Ts, the dispreference for LP Ts in its various forms (1) should be discernable in terms of statistical frequency. In fact, according to Haspelmath the PRC is just a manifestation of a yet more general phenomenon namely, the Ditransitive Topicality Role Constraint (DTRC) in (2).

(2) Grammars are likely to put restrictions on R-T combinations to the extent that the R argument is not inherently more topic worthy than the T argument.

The DTRC covers not only the dispreferred person combinations of the R and T underlying the PRC but also dispreferences involving humanness, animacy, definiteness, pronominality and cardinality. And indeed as Haspelmath documents there are various languages which preclude the T from being human, or animate, or pronominal or definite or some combination of these while simultaneously requiring the R to display precisely these properties.²

Returning to the issue of the distribution of LPs, note that since LPs are the most topic worthy participants in discourse, the various renditions of the PRC outlined in (1)) fall out of the DTRC only if Rs are taken to be preferentially LPs, in fact if they are seen to preferentially conform to the person hierarchy: 1 > 2 > 3. It is not clear whether this LP or even first person preference for Rs captured in the

² This is not to say that all the prohibitions against particular person combinations of the T and R found in languages can be automatically subsumed under the DTPR. For example, Romanian (Nevins 2007:297) is essentially a (1c) type language and thus disallows first person Ts in the presence of second person Rs. However, rather unexpectedly, it permits combinations of second person Ts and third person Rs, as in (i), which is counter to the DTRC.

(i) *Maria* i  te  a  prezentata
   *Maria* 3sg:dat  2sg:acc  has  introduced.’
   ‘Maria has introduced you to her.’

The North Italian dialect spoken in Switzerland Bellinzonese (Cattaneo 2009), by contrast, allows the very unusual combination of a first person T and second person R as in (ii) but precludes the more common second person T and first person R allowed in (1c) type languages.

(ii) *I ma  t    - à  present*
    they 1sg:acc  2sg:dat  have  presented
    ‘They introduced me to you’

I will not be going into all the reflections of the PRC in languages which have grammaticalized some form of this constraint as I am primarily interested in its frequency effects in the languages which have not grammaticalized it. However, the above patterns will be further discussed in Section 5.
DTRC should be interpreted only in relative terms, i.e., in comparison to Ts, or as a general property of Rs in ditransitive clauses. There is no reason to doubt that the former is the case. But what about the latter? Haspelmath does not comment on this issue presumably as it is not central to his thesis. Nonetheless, the question arises, are Rs prototypically LPs or even first persons?

The fact that Rs are prototypically human cannot be interpreted as automatically implying an LP preference, though this seems to be often assumed. That no such equation is possible is strongly suggested by Dahl’s (2003) investigation of the distribution of LPs in transitive and intransitive clauses in Swedish conversation. His study of a small 65 thousand-word corpus reveals that even transitive As, which are strongly associated with human pronouns occur as LPs in only 36% of the cases. This figure rises to 54% if only animate As are considered but is still definitely not overwhelming. For intransitive Ss and transitive Ps, even just animate ones, the figures are considerably lower, namely 35% and 20% respectively. Interestingly enough, although Dahl dismisses the nine instances of ditransitive clauses in his data as statistically meaningless, in all but one of these the R was an LP. Might not then Rs display a closer relationship to LPs than As, let alone Ss? Dahl’s study also reveals that the relationship between LP status and A, S and P function is strongly affected by the nature of the predicate. Thus, for example, in the case of mental predicates the percentage of LPs as As rises from 54% to a hefty 82%. This suggests that the frequency of LP Rs occurring with ditransitive predicates expressing physical transfer of possession such as give, hand or loan may well be significantly different from Rs found with predicates expressing message transfer, such as tell, teach or read, or predicates depicting spatial transfer, such as send, mail or throw. Further, since ditransitive predicates are known as often being able to occur in more than one type-of ditransitive construction, as is the case with the above cited predicates in English, a factor which may also be expected to have a bearing on the frequency of LPs as Rs is the type of ditransitive construction that it occurs in.

To the best of my knowledge the nature of the association between ditransitive Rs and LPs is not a topic that has been previously explicitly examined. This paper will do so on the basis of corpus data from two languages, English and Polish, both of which have a rich array of ditransitive predicates. The two languages differ in the primary ditransitive construction that they display, which in English is the so called double object construction (DOC), and in Polish the dative accusative construction (DAC) (see Section 2.2 for further discussion). However, both languages also have one or more prepositional constructions in which the argument seen as corresponding to the R of the primary ditransitive and which nowadays is often analyzed as a spatial goal (G) rather than an R (see e.g., Levin 1993; Goldberg 2005), despite its humanness, is prepositionally marked. The two
languages thus provide us with a good opportunity to examine the constructional effects on the frequency of LPs, as Rs, or R/Gs, both language internally and externally. We will be able to compare the distribution of LPs in positionally marked Rs (in the DOC) with that in case-marked ones (in the DAC), and both with the LP levels of prepositionally marked R/Gs.

Our investigation of LP Rs in English and Polish ditransitive clauses will also provide us with an opportunity to consider the frequency effects of the dispreference for LP Ts, i.e., how the PRC is reflected in the two languages. Recall that if the PRC is just a grammaticalized rendition of the general dispreference for LP Ts and ultimately of the DTRC as argued by Haspelmath (2004), we should find statistical reflections of the dispreferences in (1a)–(1c) in any given language, irrespective of the formal realization of its T and R pronouns. There is every reason to suspect that this is so. Moreover, in the light of the DTRC, the strength of these dispreferences should vary with the degree of grammaticalization of the person forms and potentially the ditransitive construction itself. One might therefore expect not only stronger restrictions among affixal forms than clitic ones, and among clitic forms than nonreduced ones, but also among the nonreduced forms occurring in different ditransitive constructions. Polish, unlike English, has morphologically distinct reduced person forms for the T and R. Together the two languages also display an interesting array of ditransitive constructions which may be seen as differing in degree of grammaticalization. A consideration of how the PRC is reflected across the different constructions and in the two languages should therefore constitute an interesting testing ground for some of the more detailed aspects of Haspelmath’s claims.

The presentation is structured as follows. Section 2 presents a brief overview of the most important characteristics of English and Polish ditransitive constructions and the nature of the person forms used in these constructions in the two languages. Equipped with the relevant information we then turn in Section 3 to the design of the study pointing out how it has been shaped by the findings of previous corpus research on ditransitive constructions. The relationship between LPs and ditransitive Rs is examined in Section 4. After explicating the range of ditransitive predicates taken into account in the investigation, we first consider the distribution of LP Rs in the English DOC and Prep-C and then proceed to provide the corresponding data for the Polish DAC and Prep-C. From both languages two sets of data are looked at; the first set is drawn from a mixed but predominantly written corpus, the second from a spoken corpus. In Section 5 we launch into a consideration of the frequency effects of the PRC again, first in English and then in Polish, and finally across the different types of ditransitive constructions. The discussion closes with some concluding remarks in Section 6.
2 Ditransitive constructions in English and Polish: A brief overview

As mentioned in the introduction English has two ditransitive constructions, which are referred to either via their morphosyntactic realization as the double object construction (DOC) and the prepositional construction (Prep-C) respectively, or in terms of their primary semantics as the transfer of possession and caused motion constructions (see e.g., Levin 1993; Hudson 1992; Goldberg 2005). In the DOC both the R and T are morphologically unmarked and are thus morphologically indistinguishable from each other, while in the Prep-C the T is morphologically unmarked and the R or R/G is preceded by the preposition to. The DOC is associated with R > T order, as in (3a), the Prep-C with T > R order, as in (3b).

(3) a. John gave me the book.
   b. John returned it to his brother.

There is, however, a little variation possible. In some dialects T > R order is the norm in the DOC when the R is a human pronoun and the T a nonhuman pronoun (see Gast [2007] or Siewierska and Hollmann [2007] for a discussion) as in (4a), and in most dialects R > T order is possible or even favored in the Prep-C when the T is exceptionally long or complex, as in (4b).

(4) a. He gave it me.
   b. This irregularity in her features was not grotesque, but charming and gave to Anastasia’s face a humor she herself did not possess.
   (Biber et al. 1999:928)

Some verbs occur only in the DOC construction, others only in the Prep-C, yet others can occur in either, be it in most cases with evident preferences for the DOC or the Prep-C. We will be concerned primarily with constructions featuring predicates of the last type, i.e., with the so-called alternating predicates.

While the pronominal Ts and Rs in the DOC do not differ from those used in the Prep-C in terms of their segmental phonology, there are significant differences between the two which are suggestive of the former being more strongly gram-

---

3 There is a wealth of literature on what determines the DOC/Prep-C alternation, most of which is not directly related to the current study. The major factors discussed are: grammatical relations (see e.g., Hudson 1992), length (see e.g., Hawkins 1994), information structure, (see e.g., Polinsky 1998), and animacy (see e.g., Bresnan et al. 2007).
maticized than the latter. According to the typology of structural deficiency developed by Cardinaletti and Starke (1999), pronominal forms may be divided into strong, weak and clitic forms. Only strong forms may be coordinated and modified. Weak forms or clitics cannot. Weak forms, unlike clitics, however, may bear word stress and may be deleted under ellipsis. They do not, on the other hand, form clusters. In terms of this typology, the pronominal Ts and Rs in the Prep-C emerge as strong pronouns since they can be modified and coordinated as well as separated by a verbal particle as shown in (5).

(5) a. *He gave them all to me and her.*  
   b. *He’ll send both me and you to her.*  
   c. *He gave it just to me.*  
   d. *They gave it back to me.*

The status of the pronominal Ts and Rs in the DOC is less clear. For speakers who allow only R > T order with pronominals, the R is something between a weak form and a clitic. It is typically unstressed and cannot be separated from the verb. In speech it may even fuse with the verb, as in the case with the verb *give* and *me* in *gimi it*. It can, on the other hand, be elided as in *I gave $20 in answer to How much did you give her?* The T, on the other hand, can be coordinated (6a) and modified (6b), which is not consistent with it being a weak form.

(6) a. *He gave me both her and him.*  
   b. *He gave us it all.*

Nonetheless, it is extremely rare for a pronominal R and T to be separated from each other (there are no instances in the BNC), which is suggestive of the two forming or being in the process of forming a unit. This, in turn, is not typical of a combination of a weak and strong form. Therefore for the purposes of this study I will treat both as weak forms.

In the basic ditransitive construction in Polish (DAC), the R is in the dative case and the T is in the accusative.4 The accusative and dative cases are morphologically distinct from each other in all nominal declensions. The same holds for the pronominal forms apart from the first person singular. There are two sets of

---

4 Alternative case marking patterns are also found. For example, with the verb *uczyć* ‘teach’, the R is in the accusative and the T in the genitive. See Dziwirek (2002) for the origins of this marking pattern.
pronominal forms for both the R and the T, strong forms and clitics.\(^5\) As shown in (7), the first person singular strong T (accusative) and R (dative) forms are homophonous. The strong and clitic forms are morphologically distinct only in some of the persons in the singular, the dative forms in the 1sg, 2sg, 3sgm and 3sgn, the accusative forms in the 2sg, 3sgm and 3plv.\(^6\)

\[(7) \quad \begin{array}{cccccc}
\text{Strong Forms} & \text{Clitics} \\
\text{Accusative} & \text{Dative} & \text{Accusative} & \text{Dative} \\
1\text{sg} & \text{mnie} & \text{mnie} & \text{mie (mię)} & \text{mi} \\
2\text{sg} & \text{ciebie} & \text{tobie} & \text{cię} & \text{ci} \\
3\text{sgf} & \text{ją} & \text{jej} & \text{ją} & \text{jej} \\
3\text{sgm} & \text{jego} & \text{jemu} & \text{go} & \text{mu} \\
3\text{sgn} & \text{je} & \text{jemu} & \text{je} & \text{mu} \\
1\text{pl} & \text{nas} & \text{nam} & \text{nas} & \text{nam} \\
2\text{pl} & \text{was} & \text{wam} & \text{was} & \text{wam} \\
3\text{plv} & \text{ich} & \text{im} & \text{je} & \text{im} \\
3\text{plnv} & \text{je} & \text{im} & \text{je} & \text{im} \\
\end{array}\]

The third person singular forms are used for humans, animates, and inanimates (see (8) below), though for the last of these demonstratives may also be employed.

Since Polish is a language with highly flexible word order, the T and R can occur in any order relative to the verb or each other. Thus, a ditransitive clause with an overt subject (Polish is a null subject or pro-drop language), a verb and a nominal T and R may occur in one of 24 possible orders. Clitics, which often, though by no means invariably, occur in Wackernagel’s position (cf. (8a) with (8b) and (8c)), tend to form clusters in which the dative typically precedes the accusative, as in (8).

---

\(^5\) The clitic status of the Polish pronominal object forms is not uncontroversial. Rappaport (1989:320) suggests that they are becoming less clitic like and more like orthotonic pronouns. A similar position is expressed by Franks (1998:84). In terms of the typology of Cardinaletti and Starke (1999) Polish object clitics are in between clitics and weak forms. While they are typically seen as forming clusters, these are not invariant with respect to order (see (9) below) and therefore are not considered as true clusters by some. Despite these controversies, I will continue to call the relevant forms clitics.

\(^6\) The controversies surrounding the status of the object clitics involve both the forms that are phonologically distinct from the strong forms and those that are not. It seems to be generally accepted that the latter can function either as clitics or as strong forms (see e.g., Rappaport 1989:325).

\(^7\) The clitic for the first person singular mię is essentially archaic.
Local pronouns in ditransitive scenarios

(8) a. Podała mu je do przymiarki
    handed:3SGF 3SGM:DAT 3PL:ACC during to try on
    ‘She handed them to him to try on.’

b. Sąsiedzi zostawili mi go na weekend.
    neighbors left 1SG:DAT 3SGM:ACC for weekend
    ‘The neighbors left me it/him (the dog) for the weekend.’

c. To właśnie Andrzej mi go polecił.
    Emph specifically Andrej 1SG:DAT 3SGM:ACC recommend
    ‘It was precisely Andrej who recommend him to me.’

Nonetheless, acc > dat orders also occur, as in (9). (See also Section 5.2).

(9) a. Najpierw zabierał go jej sport, później praca.
    First take away 3SGM:ACC 3SGF:DAT sport later work
    ‘At first sport took him away from her, then work.’

b. Mówią: wydajcie go nam
    say release:2SGPL 3SGM:ACC 1PL:DAT
    ‘They say, release him to us’

Discontinuous sequences of person forms are also occasionally encountered.

While the DAC is used in Polish for a wide range of trivalent scenarios, much more widely than the DOC in English, there is also a ditransitive prepositional construction (Prep-C) where the T is in the accusative and the R or rather R/G is introduced by the preposition do ‘to’ as in (10) or in the case of verbs of dispossess-sion od ‘from’ as in (11).

(10) a. Ojciec posłał mnie do niego po książkę.
    Father sent me:acc to him  for book
    ‘Father sent me to him for a book.’

b. On nigdy nie zabiera do mnie kota
    He never not take:3SG to me cat:GEN
    ‘He never takes the cat to me (my place).’

c. Powiedział do mnie „Jak śmiesz!”
    said:3SGM to me how dare:2SG
    ‘He said to me “How dare you.”.’

8 In negative clauses the T (like the P of transitive clauses) occurs in the genitive case rather than in the accusative.
(11) a. Zabieram dzieci od matek
take:1sg children:acc from mothers ‘I take children away from their mother’s.’
b. Odebrał ode mnie telefon
took:3sgm from me phone ‘He took/received the phone from me.’

As argued by Dąbrowska (1997:127–140), the Prep-C construction in Polish is a much more restricted alternative to the DAC, than the Prep-C construction is to the DOC in English. There are not many verbs which can occur in both constructions, and the meaning of the prepositional object is clearly different from that of a dative-marked one, which in Polish, as in most Slavic languages, is strongly associated with the “personal sphere”. The Prep-C in Polish will therefore play a much smaller role in our discussion than its equivalent in English. Nonetheless, since it may constitute a potential alternative to the DAC, it must be taken into account.

3 Design of the study

The design of this study owes much to the findings of previous corpus based investigations of ditransitives. How these have guided the current investigation is briefly report on in Section 3.1. The corpora used and the strategies of data retrieval employed are outlined Section 3.2.

3.1 Previous corpus investigations of ditransitives

Previous corpus based investigations of ditransitives have revealed that ditransitive clauses are considerably less frequent in texts than transitive and intransitive ones and are heavily dominated by just a few predicates. (see e.g., Gries and Stefanowitsch 2004; Mukherjee 2005; Siewierska and Hollmann 2007; Colleman 2009). For example, Mukherjee (2005:82) identified 1741 instances of ditransitive constructions in the 1 million word British part of the International Corpus of English (ICE) embracing 70 verbs. Of these, 1053 (61%) involved just two verbs give (562) and tell (491) with 24% of the 70 predicates occurring only once. In Colleman’s nine million word corpus which yielded 16.065 ditransitive constructions, the Dutch equivalent of give, namely geven was similarly dominant; 21% (3400) of the ditransitives involved geven.
Another important observation regarding ditransitives emerging from previous corpus studies is that for most if not all of the predicates found in ditransitive constructions their ditransitive use is quite clearly not their most common one. This may hold even for prototypical ditransitive predicates such as give and show. In Mukherjee’s study (2005:84), for example, only under half of the instances of give displayed what he terms “explicit ditransitive syntax”. Significantly for 62 of the 70 predicates occurring in the ditransitive constructions in his corpus, the corresponding figure was well under 5%. In the case of give a large number of the non-ditransitive uses were just cases of discourse based ellipsis. However, with many of the other predicates the non-ditransitive uses were evident manifestations of alternative senses of the predicate in question and alternative argument structures.

The final finding of previous investigations of ditransitive constructions that was of direct relevance to this study relates to the grammatical categories of the R and T. With prototypical ditransitive predicates the most common pattern (see e.g., Siewierska and Hollmann 2007 for a summary of data on English) is for the R to be a pronoun and the T a lexical NP (e.g., She told him a wonderful story.). Especially infrequent are constructions where both the T and R are pronominal, and even more so when the two are both pronominal and human and thus fall within the scope of the PRC (see Section 5).

The lesson to be drawn from the above is that if one wants to investigate the distribution of LPs in a wide range of ditransitive predicates rather than just a few statistically dominant ones, and take into account not only Rs but also Ts, one needs a very large corpus. However, since the identification of actual ditransitive uses of the various predicates will have to be done manually (there are no very large corpora which are explicitly annotated for ditransitivity), some restrictions need to be imposed to render the investigation manageable. The restrictions may pertain to the corpus or the properties of the ditransitive constructions to be considered or to both. To ensure the best possible coverage of ditransitive constructions relevant to the investigation, I adopted a two-pronged approach.

For the part of the investigation centered on the LP status of the R, irrespective of the nature of the T, I limited the number of ditransitive predicates to be considered. To provide a reasonable coverage of the different semantic subclasses of ditransitive predicates, I settled for 18 predicates in each language. The rationale for the choice of predicates is discussed in Section 4. I did not impose any restrictions on the arguments other than that they needed to be overt. For the second part of the investigation focusing on the PRC, I took into account all the ditransitive predicates in the respective corpora but only in constructions where both the T and R are personal pronouns.
3.2 Data collection

The data for English are drawn from the CQP-web version of the British National Corpus, which numbers 100 million words. The data for Polish stem from two corpora, the sampler CQP-web version of the PELCRA corpus numbering 14 million words and the IPI PAN corpus consisting of 250 million words. The PELCRA corpus was used for the first part of the investigation restricted to 18 predicates. Since there were too few instances of pronominal R and T combinations in this corpus, the second part of the investigation was carried out on the much larger IPI PAN corpus. All three corpora consist primarily of written texts but each has a spoken component involving various genres, transcripts of television and radio broadcasts, plays, parliamentary debates and some conversation. The spoken component of the BNC amounts to 10 million words. The spoken-conversational component of the PELCRA Corpus contains 2.5 million words in the form of transcriptions of spontaneous conversations as well as radio and television interviews. The IPI PAN corpus contains 37.5 million words of transcripts of parliamentary proceedings only. The investigation of the LP status of Rs will be carried out first on data drawn from the complete corpora and then the results will be compared with those stemming just from the spoken components of the respective corpora. We will thus be in a position to get an impression of how distinctive the distribution of LP Rs is in predominantly written as opposed to spoken genres. In typological studies the latter are considered to be much more relevant than the former since relatively few languages have an established written tradition and, even more importantly, grammaticalization is seen to be driven by the patterns of verbal interaction found in speech rather than in writing. Nonetheless, an appreciation of the distributional patterns found in predominantly written texts is also of interest especially for comparative and diachronic studies which are typically based solely on such texts.

Since the BNC is tagged for parts of speech and the CQP-web version allows for complex category string searches, extracting ditransitive constructions from the BNC was relatively straightforward. In doing so, only consecutive strings of

---

9 For various aspects of the three corpora the reader is referred to the following websites and publications. http://www.natcorp.ox.ac.uk; Hardie (2012); Hoffmann et al. (2008); Przepiór-kowski (2004); Waliński and Pęzik (2004). Some of the data should be considered literary and/or archaic.

10 The fact that the spoken data are not just conversational data but in many instances much more scripted reduces the value of the insights stemming from the data as a reflection of the patterns of distribution in spontaneous speech. However, this is the best approximation of what happens in speech than I am in a position to currently offer.
the R and T were catered for. The looked for argument R and T combinations were: Pro NP, NP Pro, Pro Pro, and NP NP for both the DOC and prepositional constructions. (Instances of Heavy NP shift as in (4b) were not catered for.) The pronominal arguments were extracted exhaustively. In the case of NP arguments, however, the searches were restricted to phrases consisting of one or more optional determiners, followed by one or more optional adjectives, followed by the head noun and then by one or more optional prepositional phrases. In the few instances that very large data sets were produced, the thinning option was used. After extraction, the data were subjected to manual scrutiny to eliminate all the strings other than the ditransitive ones, and for sequences of pronouns, to check the T and R assignments.

Data mining in the case of the Polish corpora was more cumbersome. The CQO-web version of PELCRA is not tagged for parts of speech and even if it were, due to the inflectional case marking and many declensions of Polish nouns, as well as the word order flexibility of Polish (see above), such tagging would be considerably less helpful than it is for English. The IPI PAN corpus, in turn, has considerable search tools built around it, but these are not widely accessible. Therefore the searches for ditransitive constructions in both corpora were lemma based and all the irrelevant material had to be eliminated manually. In the case of the PELCRA corpus, the thinning option was also used in a few cases. The searches for the R and T pronoun combinations in the IPI PAN corpus were restricted to uninterrupted sequences of the two.

4 The strength of the relationship between Rs and LPs

For the practical reasons outlined in 3.1, the investigation of the affinity between Rs and LPs is confined to a representative subset of 18 ditransitive predicates from each of the two languages. Therefore before presenting my findings with respect to the LP distribution of Rs, let me first say a few words about my choice of predicates.

11 In most varieties of English the NP Pro combination is precluded from the DOC, a phenomenon which is sometimes referred to as Larson’s (1988) constraint. However, there are exceptions as noted by Gast (2007) and also Siewierska and Hollmann (2007).

12 In the case of ask and tell in English and powiedzieć in Polish clausal arguments were also taken into account.
4.1 Choice of predicates

In choosing the ditransitive predicates for the study I sought to achieve a good coverage, on the one hand, of the different semantic subclasses of these predicates and, on the other, of the most frequently occurring predicates.

English ditransitive predicates are grouped by Goldberg (2005:126) into the nine subclasses in (12).

(12) a. Verbs of giving: give, hand, lend, lease, rent
   b. Verbs of instantaneous causation of ballistic motion: fling, throw, toss
   c. Verbs of sending: send, mail, forward
   d. Verbs of continuous causation of accompanied motion: bring, take
   e. Verbs of future having: offer, promise, assign
   f. Verbs of communicated message: tell, show, ask, teach, read
   g. Verbs of instrument of communication: e-mail, telegraph, wire
   h. Verbs of creation: bake, make, build, cook, sew
   i. Verbs of obtaining: get, buy, find, steal, win

The lexical scope of Polish ditransitive predicates is wider than in English (Rudzka-Ostyn 1996). In addition to the nine subclasses in (12), there are predicates of dispossession (e.g., zabrać `take away’, odebrać `take back’, ukrąć `steal’), enabling (e.g., udostępnić `make available’, ustąpić `yield’), hindrance (e.g., zabronić `forbid’, zakazać `prohibit’), causation (e.g., uwidocznić `make visible/explicate’, uzmysłowić `make apparent in one’s mind’, sprawić `cause’), and imposition (e.g., podporządkować `subordinate’, poświęcić `devote’). However, to maintain parity of comparison, I restricted the choice of predicates to the subclasses that the two languages have in common.

Nonetheless, not all of the subclasses in (12) are represented in the 18 predicates that I finally selected for the investigation. Verbs of class (g), i.e., of instrument of communication are rarely used in Polish as the language prefers to express such meanings with a species of a send verb and an NP (e.g., wysłać tele- gram, wysłać e-mail). Therefore this subclass was disregarded. I also disregard the predicates from the last two subclasses in (12h) and (12i), the predicates of creation and obtaining, but for quite different reasons. In both languages these predicate are well represented in ditransitive constructions. However, the semantic role of the non-agentive human argument in the DOC and DAC found with these subclasses of verbs is between that of recipient and beneficiary, a recipient-beneficiary in the terminology of Kittilä (2005). This is reflected in the use of the preposition for rather than to in English and dla rather than do in Polish in the corresponding prepositional constructions, as shown in (13).
Local pronouns in ditransitive scenarios

(13) a. Wybudował/ znalazł  mi  dom.
built:3sgm/found:3sgm 1sg:dat  house:acc
‘He build/found me a house.’
b. Wybudował/ znalazł  dom  dla mnie/* do mnie.
built:3sgm/found:3sgm  house:acc  for me  to me
‘He built/found a house for me.’

Although the distinction between recipients and beneficiaries is by no means a clear cut one (see especially Kittilä 2005), in order to remain as close as possible within the sphere of prototypical recipients, these subclasses of ditransitive verbs were also not taken into account.

This leaves us with the top six subclasses in (12). For English, I selected three verbs of giving, ballistic motion and sending, two of accompanied motion and future having and five of communication of message13. For Polish I did the same with one exception. Polish, unlike English, has only one predicate in the ballistic motion subclass that is regularly used, namely rzucić ‘throw’. Therefore to have an equal number of predicates from both languages rather than adding some from one of the other subclasses, I added two predicates of dispossession which are used very frequently in Polish. The final list of predicates selected for each of the two languages is presented in (14).

(14) a. English: (a) give, hand, lend (b) fling, toss, throw (c) forward, mail, send (d) bring, take (e) offer, promise (f) ask, read show, teach, tell14
b. Polish: (a) dać, wręczyć, podać (b) rzucić (c) przesłać, przysłać, wysłać (d) przynieść, zamieść (e) obiecać, oferować (f) pokazać, powiedzieć, przekazać, przyznać, uczyć (g) zabrać, odebrać

It is worth noting that the English list contains all the five predicates which on the basis of their frequency of occurrence in the ICI corpus were classified by Mukherjee (2005:82) as being “typical ditransitive verbs” (give and tell) and “habitual ditransitive verbs” (show, ask, send, offer). I am not aware of any comparable frequency counts of Polish ditransitive verbs which I could cite in this context other than my own which, however, are restricted to ditransitive clauses with pronominal Ts and Rs. These suggest that the list of 18 Polish ditransitive predicates selected for this investigation also features the most frequently occur-

13 For alternative subclassifications see Pinker (1989); Levin (1993); Croft (2001).
14 This potential overrepresentation of predicates of communicated message is due to my wanting to test whether these predicates would display a higher level of LP Rs than the other types, as did mental predicates with respect to the A in Dahl’s (2003) study mentioned earlier.
ring ditransitives in Polish, namely dać, odebrać, przekazać, zabrać, pokazać, podać, wręczyć and przyznać.

4.2 LPs as Rs in English

All the predicates listed in (14a) occur in ditransitive constructions in the full BNC though not in the spoken part of the corpus. We will therefore begin the discussion with the data stemming from the full corpus and then compare the results with those originating just from the spoken part of the corpus in Section 4.4.

Using the procedures outlined in Section 3.2 I identified in the BNC 11,514 instances of ditransitive constructions featuring the 18 predicates listed in (14a). The numerical data are in Table 1 in the Appendix. Among these 11,514 ditransitive clauses, only 35% (3,863) displayed a LP R. This low figure suggests that Rs in ditransitive clauses are not typically LPs in English, at least not in ditransitive clauses in general, i.e., undifferentiated for predicate class and/or type of construction. When we consider the six subclasses of predicates a more varied picture emerges. Nonetheless, only for one subclass of predicates does the frequency of LPs as Rs rise to over 50%; with predicates of communicated message, 52% of the Rs are LPs. Interestingly the corresponding figure for giving predicates is only 27%. The frequency of LP Rs with ballistic predicates is especially low, namely 7%. As one might expect, the frequency of LP Rs in constructions with the six subclasses of ditransitive predicates correlates with their respective propensity to be human, and also pronominal. This is depicted in Figure 1.

![Diagram](image)

**Fig. 1:** LP Rs, human Rs, and pronominal Rs in six subclasses of ditransitive predicates

---

15 In line with Goldberg (2005) and others I have grouped *bring* and *take* together. However, in the BNC data the two verbs behave very differently in regard to the ditransitive constructions that they favor and their use of LP Rs. In the DOC the LP level of the R for *bring* is the second highest (63%), while *take* has the lowest percentage (5%).
Note, however, that there is a huge difference between humanness and LP status of the R in all the predicate classes but for the accompanied motion one in which the humanness of the R is exceptionally low. With giving verbs, for example, there are more than three times as many human Rs than LP ones. Thus, contrary to what is often assumed, humanness alone cannot be taken as indicative of the likelihood of an R being an LP. Pronominality fares somewhat better as a guide to LP status than humanness, especially with sending predicates. Nonetheless, as Figure 1 demonstrates, there is still quite a gap between the levels of the two.

So far we have been considering the distribution of LPs among Rs irrespective of the type of ditransitive construction that they occur in, DOC or Prep-C. Since the latter may be viewed as containing Gs rather than Rs, before jumping to any conclusions with respect to the relationship between LPs and Rs in ditransitives, it is crucial that we distinguish between the two constructions.

Given that the primary ditransitive construction in English is the DOC, it should come as no surprise that among the 11,514 instances of ditransitive clauses constituting the English sample of ditransitives, there are nearly twice as many DOCs (7,354) as Prep-Cs (4,160). The distribution of LPs as Rs in the two constructions is quite distinct; whereas nearly half, 48% (3,550) of the DOCs have an LP R, only 8% (313) of the Prep-Cs do so. That LP Rs should favor the DOC over the Prep-C, even so strongly, is hardly unexpected. What is a surprise though is that the level of LP Rs in DOCs is still under 50%.

When we consider the distribution of LP Rs relative to predicate subclass we find that the strong preference for LP Rs in the DOC as compared to the Prep-C is reflected in each group of predicates. This is depicted in Figure 2.

Figure 2 also reveals that in three of the six subclasses the level of LP Rs in the DOC rises above 50%, be it not dramatically so; the highest level of LPs is in the

Fig. 2: LP Rs in the DOC and Prep-C relative to predicate subclass
accompanied motion subclass and it is just 60%. Interestingly, verbs of giving score only 37%, which is the second lowest score after the ballistic verbs (13%). This is important as verbs of giving and especially *give*, which has a score of 42%, are typically considered as the most prototypical of ditransitive verbs.  

While one might be tempted to quibble as to whether levels of LP Rs of around 60% should be interpreted as indicative of the dominance of LP Rs at least with certain subclasses of ditransitive predicates in the DOC or not, it is more instructive to compare the LP realization of Rs with their human and pronominal status, as we did in the case of the DOC and Prep-C earlier in Figure 1. The data in question are presented in Figure 3.

We see that for all the predicate subclasses the level of human Rs is or is near categorical and of pronominal Rs does not fall under 82%. Human Rs and even pronominal ones are thus clearly prototypical for the English DOC. By contrast, LP Rs cannot be thus considered, without rendering the notion of prototypicality vacuous.  

### 4.3 LPs as Rs in Polish

From the Polish PELCRA corpus I identified 4,048 instances of ditransitives comprising both DAC and Prep-Cs with the 18 predicates listed in (14b). The numerical data are presented in Table 2 in the Appendix. The percentage of LP Rs among

---

16 The instances of *give* were capped at 5%, and of *send* and *show* at 20%.
17 This is not always so. Kittilä (2006) discusses the peculiarities of *give* in many languages.
18 LP Rs emerge as dominant only if we confine our attention to pronominal realizations of the R, 56% of the pronominal Rs in the corpus are LPs.
these 4,048 ditransitive constructions is slightly lower than in the corresponding English sample, namely 30% (1,232). Also somewhat different is the distribution of LPs in some of the subclasses of the ditransitive predicates. As shown in Figure 4, the level of LP Rs in Polish is quite close to that of English in verbs of giving, communication, accompanied motion and ballistic. Verbs of sending and future having, by contrast, diverge quite substantially, both manifesting twice or three times as few LP Rs in Polish as in English. The predicates displaying the highest level of LP Rs are the same in both languages, namely the communicated message ones, though the level of LP Rs in Polish is lower (47%) than in English (52%). There is thus no subclass of ditransitive predicates in Polish for which the level of LP Rs exceeds 50%.

The situation with respect to the occurrence of LP Rs hardly changes if we separate out the DACs from the Prep-Cs. As in English, the difference in the frequency of LP Rs in the primary ditransitive construction, DAC in Polish, as compared to the Prep-C is enormous, namely 37% vs. 3%. Unlike in English, in none of the subclasses of predicates does the level of LP Rs in the primary ditransitive construction exceed 50%. The highest scoring predicates for LP Rs within the DAC are the verbs of communicated message which score 49%. There is again a very substantial gap between these and the subclass of predicates with the next highest score in regard to the level of LP Rs, verbs of sending (33%), while verbs of

---

19 The level of human and pronominal Rs in the Polish ditransitive constructions is also lower being 67% and 52% respectively.
20 The distribution of LP Rs among the verbs of communicated message is, however, quite different in Polish than in English. In English, it oscillates between 64% for *tell* and 43% for *show*, whereas in Polish the top scoring *powiedzieć* ‘tell’ scores 65% and the bottom scoring *przyznać* ‘attribute’ 0%.
accompanied movement, which score the highest 60% in English, score only 25% in the Polish DAC. The case for treating LPs as prototypical Rs with any of the verbal subclasses of ditransitive predicates in the DAC in Polish is thus even less compelling than with respect to the English DOC. The relevant data from the Polish DAC together with that from the English DOC are presented in Figure 5.

4.3 LP Rs in the DOC, DAC and Prep-C in the spoken corpora

The corpus data that we have been considering have shown that in both English and Polish there is an enormous difference in the LP level of the R in the primary ditransitive construction the DOC/DAC as compared to that of the R/G in the Prep-Cs of each of the two languages. There is thus no doubt that LP Rs favor the DOC and DAC as opposed to the Prep-Cs. Nonetheless, in neither the DOC nor the DAC was the level of LP Rs high enough to support the claim that LPs are prototypical of either of the two ditransitive constructions.

The data from the spoken sections of the English and Polish corpora suggest that in speech the level of LP Rs is significantly higher than what might be expected from a consideration of a mixture of written and spoken texts. Nonetheless, while the differences are statistically significant (p < .005%), in neither language are they of an order of magnitude which one might be sometimes led to expect. Let us first consider the English data.

The spoken part of the BNC yielded 5,128 ditransitive clauses with 14 of the 18 predicates in (14a). There were no ditransitive constructions in the corpus with toss, fling, mail and forward and only negligible instances of those with throw (7) and promise (9). The data are heavily dominated by predicates of giving and communicated message. Especially dominant is give itself, which is the predicate in

![Fig. 5: LP Rs in the Polish DAC as compared to the English DOC relative to predicate subclass](image-url)
69% of the ditransitive clauses of the spoken corpus. A most notable difference relative to the full corpus is a substantial increase in pronominal Rs, from 61% in the full corpus to 90% in the spoken part. Given the nature of the relationship between pronominality and LP status depicted earlier in Figure 1, we may expect the increase in pronominality to be accompanied by an increase in LP status of the R. And this is indeed so. Among these clauses 62% manifest an LP R, which is nearly twice as many as among the data from the full corpus (34%).

The distribution of DOCs as compared to Prep-Cs in the spoken data is rather different than in the full corpus. Whereas in the full corpus the ratio of DOCs to Prep-Cs was less than 2:1, in the spoken data DOCs dominate by over 5:1, the relevant figures being 4,315 vs. 813. Interestingly the overwhelming dominance of the DOC over the Prep-C in the spoken corpus is not echoed by a parallel increase in the frequency of LP Rs. Although the level of LP Rs in the DOC is considerably higher in the spoken corpus than in the full corpus (68% vs. 48%), the level of LP Rs in Prep-Cs increases much more dramatically, namely from 8% in the full corpus to 38% in the spoken one. This too correlates with a much greater increase in the pronominality of the R in the Prep-C (from 15% to 51%) as compared to the DO (from 56% to 97%).

As for the distribution of LP Rs in the different subclasses of predicates, in all subclasses but for the ballistic, which were disregarded due to their virtual absence in the spoken corpus, the level of LP Rs is considerably higher. What is especially worthy of note is the enormous increase in the level of LP Rs with predicates of giving in the DOC, from 37% in the full corpus to 65% in the spoken data. In the case of the DOC the highest scoring predicates are again the communicated message, with a score of 78%, and the lowest the accompanied motion with a score of 63%. In the Prep-C the communicated message predicates also manifest a surprisingly high level of LP Rs, namely 67% while the accompanied motion ones also score the lowest, with a very low 5%. The just mentioned data from the DOC and Prep-C are depicted in Figure 6.

Turning to the Polish data, from the spoken sections of the PELCRA corpus I extracted 1068 ditransitive constructions. As in the case of English, the Polish spoken corpus is heavily dominated by predicates of giving (291) and communicated message (547), though, unlike in English, with the latter clearly prevailing. Again the level of LP Rs in the Polish spoken data is significantly higher than in the full corpus, 51% vs. 30%. This too is partially attributable to an increase in the pronominality of Rs (from 52% to 62%) although the difference is much smaller

21 Recall that in the full corpus the thinning option was used for give and the instances of give were capped at 5%. Accordingly, ditransitive clauses with give account for only 13% of the data in the full corpus.
than in the corresponding English data. In contrast to what we saw in English, the higher level of LP Rs in the Polish spoken corpus is due virtually exclusively to the primary ditransitive construction, the DAC as opposed to the Prep-C.\textsuperscript{22} Among the DACs the level of LP Rs in the spoken section of the corpus is 58% as compared to 37% in the full corpus. LP Rs therefore do constitute a clear majority in the DAC but, as in the English DOC, by no means an overwhelming one.

In view of the skewed distribution of predicates in the spoken corpus it is only the predicates of giving and communicated message which can be meaningfully commented on. The distribution of LP Rs with the former, as in the case of English, is very much higher in the spoken corpus than in the full one (62% vs. 26%). In the case of the communicated message predicates, the difference between the two sets of data is much smaller (58% vs. 49%).

We have seen that in both English and Polish the levels of LP Rs in the spoken corpora are significantly higher than in the mixed corpora. One might speculate that the levels of LP Rs would be higher still if we had had a spoken corpus of pure conversational data rather than ones containing a variety of spoken genres. However as no such corpora are available, we must satisfy ourselves with the data that we do have. These data allow us to conclude that while in spoken texts, unlike in primarily written ones, Rs in ditransitive clauses are most often LPs, the association between the two even in spoken texts is not as strong as is often suggested. It is clearly considerably weaker than that between Rs and pronominality. This holds for both languages, though especially so for the English DOC; whereas 97%...
of the Rs in the DOC originating for the spoken corpus are pronominal, only 68% are LPs. The corresponding figures for the Polish DAC are 68% vs. 58%.

The above notwithstanding, quite evidently strong associations between ditransitive Rs and LPs have emerged for subclasses of ditransitive predicates and especially for individual predicates. Predicates of communicated message score considerably higher in regard to level of LP Rs than do all the other subclasses of ditransitive predicates in both written and spoken corpora, and, interestingly, predicates of giving evince substantially higher LP R levels in speech than in writing. This is especially important in view of the dominance of such predicates and especially give and dać in speech. It is also interesting to note that in the spoken corpus some of the individual predicates have LP R levels of over 80%, namely read (88%), bring (87%), tell (82%), and in Polish przysłać (86%). The two most common ditransitive predicates in the two languages give and dać display very similar LP R levels, 66% vs. 64%.

Being very much aware of the various differences between the BNC and the part of the PELCRA corpus that I have used here both in relation to size and composition, I am rather hesitant to enter into any direct comparisons between ditransitive constructions in English and Polish. Nonetheless, it is worth noting that the difference in the level of LP Rs between the DOC and DAC is statistically significant (p < .005%) both in the data originating from the full corpus and from just the spoken corpus. The same holds for the pronominality and humanness levels of the R in the two constructions. In short, the Rs in the DOC are more likely to be LPs, pronominal and human than the ones in the DAC. Thus in terms of these three dimensions the three types of ditransitive constructions that we have been considering can be grouped with respect to the decreasing likelihood of the R being an LP, pronominal and human as in (15):

(15) DOC > DAC > Prep-C

If (15) is confirmed by data stemming from other languages, the level of LP Rs may be seen as another reflection of the degree of grammaticalization of ditransitive constructions alongside the effects of the PRC to which we now turn.

5 The reflections of the Person Role Constraint

Recall that if what underlies the constraints on LP Ts as reflected in the PRC in (1) is a general dispreference against R-T combinations in which the R is less topical than the T, as captured by Haspelmath (2004) in the DTRC in (2), all languages should display frequency effects of the PRC. Further, these effects should vary
with the degree of grammaticalization of the person forms and constructions involved, the more grammaticalized forms/constructions exhibiting stronger effects than the less grammaticalized ones. In English although the person forms of the T and R in the DOC do not differ morphologically from those in the Prep-C, the former are considered to be more grammaticalized than the latter (see 2.1.). We may therefore expect stronger effects of the PRC in the DOC than in the Prep-C. In Polish there are both strong and clitic variants of the T and R, and both may occur in the DAC (see 2.2). We should therefore find more evidence of the PRC in connection with the clitic combinations than in regard to those involving the strong forms. Moreover, as in English, the Polish Prep-C may be expected to be more accommodating of the dispreferred person combinations than the DAC. Let us now consider whether this is indeed so.

### 5.1 The PRC in English

As mentioned in Section 3.1, in order to test the frequency effects of the PRC in English I extracted from the BNC all the instances of ditransitive clauses in which both the T and R are realized by a personal pronoun. There were 2309 instances of such clauses in total. Of these 208 (8%) were DOCs, 2101 (92%) Prep-Cs. This overwhelming preference for Prep-Cs when both the T and R are pronominal is well known. What is interesting in the context of our investigation is that among the 208 DOCs there were only two in which the T and R were not only pronominal but also human. Significantly in both instances the T was a third person, not an LP, as (16) shows.

(16) a. *I couldn’t forgive you him.*
    b. *I’ll show you her anyway.*

In the Prep-C construction LP Ts were also rare; there were only 158 instances which constitute a mere 7% of the Prep-Cs. Interestingly just under half (75) of the LP Ts occurred with just two predicates, namely *leave* (52) and *beat* (23). All the relevant instances with *beat* co-occur with *to it*, as in (17), and form an idiomatic expression with no ditransitive semantics.

(17) a. *I’ll beat you to it.*
    b. *He beat me to it.*

---

23 It is interesting to note that all the DOCs but one are from the spoken part of the BNC.
Most of the instances of *leave* are similarly idiomatic, as exemplified in (18).

(18) a. *He just left me to it.*
    b. *I felt like leaving you to him.*

Thus, if we disregard the cases of LP Ts with *beat* and the idiomatic ones with *leave*, the number of LP Ts in Prep-Cs drops to 98, which is just under 5%.

Of central relevance to the PRC are, however, instances of T LPs when the R is prototypical in being human. Among the 2,101 Prep-Cs there were only 279 ditransitive clauses with a human T and R. Among these 60 (22%) featured LP Ts. Some cases in point are given in (19) and (20).

(19) a. *Please take me to her.*
    b. *Please remember me to her.*
    c. *I hope he would lead us to him.*
    d. *But I can introduce you to him.*

(20) a. *Just remember that your contract commits you to us for four years.*
    b. *I don’t know what reason brought you to me.*
    c. *The commander has assigned me to you.*
    d. *They’ll bring me to you in just a minute.*

Among this subset of clauses with LP Ts, the mixed scenarios, as in (19), where the R/G is a third person rather than a local one, are more common than the local scenarios, where both the T and R/G are LPs, as in (20). The relevant figures are 39 vs. 21. The mixed scenarios all run counter to the DTPC since given the hierarchy of 1 > 2 > 3 the T is higher in topicality than the R/G. Among the local scenarios, however, only the nine inverse ones involving a first person T and a second person R/G, as in (19c)–(19d), do so. Thus in all there were 48 (39+9) examples (16%) which defy the DTRC.

If we take as the reference point for the PRC only the clauses with pronominal and human Ts and Rs, rather than the whole set of ditransitive clauses of a particular type, then the 48 examples in violation of the PRC constitute 16% of the data. Whether this figure of 16% should be seen as strikingly low is difficult to assess. Suffice it to say, that the scenarios which violate the PRC are indeed infrequent in discourse even in constructions which can readily accommodate them. What is particularly significant is that there are no PRC violations in the DOC. Thus, all the examples which are in defiance of the PRC involve the less grammaticalized person forms of the Prep-C as opposed to the more grammaticalized ones of the DOC, which is what Haspelmath’s interpretation of the PRC predicts.
5.2 The PRC in Polish

My investigation of the frequency effects of the PRC in Polish was carried out on the basis of data extracted from the IPI PAN corpus consisting of 250 million words. As in the case of English, I extracted from this corpus all the instances of ditransitive constructions featuring a pronominal T and R. The search was conducted using all the possible pronominal T and R combinations as a cue and then eliminating the non-ditransitive instances. In the case of the DAC I used only sequences of dative and accusative forms (in either order) and thus ignored negative constructions which feature a genitive instead of an accusative, and atypically marked ditransitives such as those found with uczyć ‘teach’ or pozbywać się ‘deny’, which require an accusative and genitive. For the Prep-C I used only combinations of an accusative followed by either one of just two prepositions do ‘to’ and od ‘from’ which in turn take a pronoun in the genitive. Only examples with the relevant semantics were taken into account. In all, I extracted from the corpus 1,836 ditransitive clauses with pronominal Ts and Rs.

Given the enormous size of the IPI PAN corpus, the small number of ditransitives with two pronouns is rather surprising, especially in comparison to the number of examples extracted from the much smaller BNC. There are three reasons for this. The first is that Polish makes far greater use of demonstrative pronouns for inanimates than English does and constructions with personal and demonstrative pronoun were not included. The second is that Polish is an object-drop language and thus allows anaphoric Ts to be omitted under various circumstances, as is the case in (21), for example.

(21) a. Wypisała czek i dała Markowi
   wrote:3SGF check and gave:3SGF Mark:DAT
   ‘She wrote the check and gave (it) to Mark.’

b. O co chodzi w tym liście? Czy możesz mi wytlumaczyć.
   about what goes in this letter Q can:2SG I:DAT explain:INF
   ‘What is this letter about. Can you explain (it) to me?’

24 There were rather a large number of dative accusative clitic clusters which were not ditransitive, due to the fact that Polish makes wide use of the so-called ethical dative and the dative sympaticus. Moreover, sine clitics tend to cluster in Polish, some of the clitic sequences involved inter- clausal material.
The third reason for the small number of sequences of pronominal Rs and Ts is that polite reference to second persons in Polish is made via the use of the nouns *Pan/Pani* ‘lord/lady’ rather than second person pronouns. All of this results in far fewer T and R sequences with a pronominal T than one would otherwise expect.

Among the 1,836 ditransitive clauses, the overwhelming majority, 1,676 (91%) were DACs. Unlike the English DOCs, the Polish DACs do display LP Ts, though not very many, namely only 30, which is 1.8%. The DACs in which both the T and R are pronominal and human were few in number, 186 (11%). Interestingly, of the 30 examples with LP Ts, just over two thirds (21) involved local scenarios, i.e., cases were both the T and R are LPs, as in (22).25

(22) a. *Nikt* mi cię nie odbierze.
   no one 1sg:dat 2sg:acc not take way
   ‘No one will take you away from me.’

b. *Cezar* mi cię obiecał
   Cezar 1sg:dat 2sg:acc promised
   ‘Cesar promised you to me.’

c. *Los* zesłał mi ciebie
   fate sent 1sg:dat 2sg:acc
   ‘Fate sent you to me.’

d. *Nikt* mi ciebie nie ukradnie
   no one 1sg:dat 2sg:acc not steal
   ‘No one will steal you from me.’

This is quite the opposite situation to that obtaining in the English Prep-C data. All of these 21 local scenarios are direct ones, i.e., the T is a second person and the R a first person, rather than the other way round, which, if we assume the hierarchy of persons to be 1 > 2 > 3, is in fact in line with the DTRC. Thus, de facto among the 30 examples with a LP T there were again only nine which run fully counter to the DTRC in that the R is lower in topicality than the T. Of these nine examples, five have a first person T, as in (23a)–(23b) and four a second person one, as in (23c)–(23d).

(23) a. *Mu* mnie polecił
   3sgf:dat 1sg:acc recommend:3sgm
   ‘He recommended me to him.’

25 All the clauses with LP Ts also had human Rs, unlike in the English Prep-C construction discussed above.
b. *Odebrał-by mnie im znowu*
take away:3SGM-POT 1SG:ACC 3PL:DAT again
‘He would have taken me away from them again.’

c. *Dali mu ciebie*
gave:3PL 3SGM:DAT 2SG:ACC
‘They gave you to him.’

d. *Cię im zostawię*
leave:1SG 2SG:ACC 3PL:DAT
‘I will leave you to them.’

There are two further points that are worth noting in regard to the LP Ts, both in local scenarios such as (22), and mixed ones as in (23). The first concerns the form of the T and the second its location relative to the R. We see in (22) and (23) that whereas the R consistently occurs in its clitic form, the T may be a clitic, as in (22a), (22b), and (23d), or a full form (22c), (22d), and (23c). In the local scenarios the clitic seems to be preferred to the full form (12 vs. 9). In the mixed scenario, however, there seems to be a preference for the full form (7) vs. (2). If the mixed scenarios are seen as constituting a stronger violation of the PRC than the local scenarios, we would expect the less grammaticalized forms to be favored in the mixed scenarios over the local ones. And this does indeed appear to be so. As for order, in the local scenarios the order of the T and R is consistently R > T, irrespective of the form of the T. In the mixed scenarios, on the other hand, the order of the T and R in 6/9 instances is according to the person hierarchy with the first or second person T preceding the third person R, as in (22b) and (22d). In Polish, unlike in Czech (Sturgeon et al. 2010), T > R order in the DAC is not a strategy used solely in cases of PRC violations (see the examples in (9) given earlier.). However, it is possible that such order is more likely in scenarios that run counter to the DPRC.

Turning to the Prep-Cs with *do* ‘to’ and *od* ‘from’, among the 160 examples with a pronominal T and R/G there were 83 in which both of the pronouns were human. The T was an LP in 35 instances (42%). Unlike in the case of the DAC, and echoing the situation in the English Prep-Cs, most of the examples (25) involved mixed scenarios, rather than local ones, i.e., where the T is an LP but the R/G is not, as in (24).

(24) a. *Zaprowadził mnie do niego nasz kolega*
lead over:3SGM 1SG:ACC to him our friend
‘Our friend leads me over to him.’
b. \( \text{Zabiorę cię do niego} \)
   \[ \text{take:1sg 2sg:acc to him} \]
   ‘I will take you to him.’

Interestingly, among the local scenarios, there were two instances of a scenario which was absent in the DAC, namely the inverse one, where the T is a first person and the R/G a second person, as shown in (25).

(25) a. \( \text{Potem odrywa mnie od ciebie praca zawodowa.} \)
   \[ \text{later tear:3sg 1sg:acc from 2sg:acc work professional} \]
   ‘Then my professional work tears me away from you.’

b. \( \text{Ursus przysłał mnie do ciebie} \)
   \[ \text{Ursus sent:3sg 1sg:acc to 2sg:acc} \]
   ‘Ursus sent me to you.’

These two instances plus the 25 in mixed scenarios add up to 27 (33%) direct violations of the DTRC.

Thus just as in the case of English, whereas the more grammaticalized DAC is strongly in line with the DTRC with only 4% exceptions, the less grammaticalized Prep-C is much more tolerant exhibiting 33% violations.

5.3 DOC vs. DAC vs. Prep-C

We have seen that there are very clear frequency effects of the PRC in both English and Polish. Since the English DOC is intolerant of pronominal Ts and Rs and especially human ones, it is also the most restrictive in regard to the person combination of the R and T in that it prohibits any combination involving an LP T. The Polish DAC, by contrast, is rather liberal in that it disfavors or perhaps even precludes only local inverse combinations of 1 > 2, i.e., where the T is first person and the R second person. Such local inverse combinations are the most strongly disfavored in both languages. In terms of frequency of occurrence, in both languages the mixed combinations (1/2 > 3) are more common than the direct local ones (2 > 1). However, in Polish the two exhibit somewhat different distributions with respect to the DAC and Prep-C. Whereas the mixed combinations clearly favor the Prep-C, the direct local combinations do not and may even be seen as exhibiting a slight preference for the DAC. These patterns are summarized in Figure 7.

The above data are strongly supportive of there being a close relationship between type of T and R encoding and tolerance of PRC violations. We have seen that the English DOC, in which the T and R are distinguishable from each other
only by means of order, is clearly inhospitable to any combination of a T and R involving human pronouns let alone LP Ts. The Polish DAC, in which the T and R are distinguished in terms of case, be it as clitics or strong forms or a combination of the two, is relatively tolerant of local scenarios, especially the direct ones but considerably less accommodating towards mixed scenarios. These, in turn, feature rather prominently in the Prep-Cs of both English and Polish in which the T and R/G are distinguished from each other by means of the prepositional marking of the R/G.

That atypical ditransitive scenarios should disfavor the DOCs is not surprising, but the possibility that local ones might prefer the DAC to a Prep-C is. Haspelmath (2004) argues that scenarios involving what he terms reverse person associations for the T and R, i.e., the mixed scenarios in the terminology used in this paper, are more likely to occur in complex constructions, such as the Prep-C, than scenarios in which both the T and R display person values normally associated with their respective roles. However, no mention is made of scenarios in which only either the T or the R bears an atypical person role, as is the case most clearly in local scenarios in which the R continues to bear what for Haspelmath is its prototypical role, i.e., an LP one. Significantly, the preference of local scenarios for the DAC is not attributable to their greater frequency of occurrence relative to the mixed scenarios since as Figure 7 shows the latter are considerably more frequent than the former.

This also holds for the German data drawn from the COSMAS Goethe corpus that Haspelmath cites, in which mixed scenarios outnumber local ones by more than 2:1 (6% vs.15%). Perhaps the relative frequency of local scenarios in the DAC as compared to the Prep-C in Polish is simply an artifact of the data or alternatively of the overall preference for the DAC over the Prep-C in Polish. What argues against this is that a comparable distinction, be it in the form of an absolute constraint and not just a mere statistical effect, is to be found in languages which

---

Fig. 7: LP Ts in mixed, inverse and direct combinations in English and Polish

---

Unauthenticated
Download Date | 6/21/19 5:40 AM
display the weak (1b) and (1c) as opposed to strong version (1a) of the PRC such as Italian, Spanish and Catalan. In these languages, while combinations of clitics in at least one of the local scenarios are possible, those involving mixed scenarios require a Prep-C or, in some instances, a locative as opposed to dative clitic. A possible way of accounting for the above is by a modification of the DTRC, i.e., with reference to the extent of the difference in inherent topicality between the T and R. If it is assumed that the difference in inherent topicality between LPs is smaller than between either of the LPs and a third person, the discussed favoring of DACs as opposed to Prep-Cs by local scenarios could be captured in a formulation such as (26).

(26) The less inherently topic worthy the R is relative to the T, the more likely there are to be restrictions on the expression of the two and the more complex the constructions in which they may co-occur.

While I am rather skeptical of speaker’s ability to make graded distinctions such as those assumed in (26), I currently have no better explanation for the discussed data to offer.

6 Concluding remarks

This corpus based investigation of the distribution of LPs as Rs and Ts in ditransitive constructions in English and Polish has revealed that the distribution of LPs is strongly affected by the degree of grammaticalization of a ditransitive construction. Both the levels of LP Rs discussed in Section 4 and the frequency effects of the PRC considered in Section 5 have been strongest in the DOC and weakest in the Prep-Cs as shown in the cline in (15) repeated in (27).

(27) DOC > DAC > Prep-C

While this was not unexpected in the case of the PRC effects, it did come rather as a surprise as far as LP Rs are concerned. To the best of my knowledge no previous observations as to the frequency of LP Rs in DOCs as compared to DACs had ever been made.

With respect to the levels of LP Rs, we have also seen that these may differ significantly in different language modalities, i.e., written and spoken texts. That such might be the case is often assumed but again has never been documented in any detail. The data stemming from the BNC and PELCRA corpora have shown that written texts do not exhibit an outright preference for LP Rs while spoken
ones do, be it not always as strong a preference as is often supposed. Further, the levels of LP Rs are heavily dependent on the semantic subclass of the predicate, or even on the individual predicate. As in Dahl’s (2003) study, the mentally oriented predicates stood out in their preference for LP Rs. Here the relevant predicates were the communicated message ones which displayed as a group LP R levels as high as 78% in the case of the DOC in the spoken part of the BNC, and even 88% for the individual predicate read. Ballistic predicates, by contrast, were shown to display very low levels of LP Rs as did also the accompanied motion ones especially take. The prototypical ditransitive predicates of giving were found to be especially sensitive to the written vs. spoken dimension, with give and its Polish counterpart dać exhibiting very large increases in LP levels of the R in the spoken corpora as compared to the predominantly written ones.

The frequency effects of the PRC were shown to be strong both in English and Polish and in the main in line with the topicality constraints captured in the DTRC. However, the apparent preference of local scenarios, both for the DAC as opposed to the Prep-C in Polish as well as in languages manifesting the weak version of the PRC were argued not to be directly derivable from the DTRC in the form proposed by Haspelmath (2004). An alternative formulation was proposed which, however, appeals to speakers’ ability to carry out gradient as opposed to binary topicality judgments, something for which there is little evidence.

The study of the distribution of LPs in ditransitive constructions has proved to be much more revealing about the nature of ditransitive constructions than could possibly have been supposed. The differences in regard to the distribution of LP Rs and Ts between the semantic subclasses of predicates are of special interest and deserve further examination as they may underlie the emergence of alternative schemas to that defined by give and different paths of diachronic development. It is hoped that the presented data will encourage more research on the topic.

Acknowledgments: The research for this paper is part of the work of IP4 (Hierarchical ranking and argument encoding in three participant clauses) of the Referential Hierarchies in Morphosyntax CPR carried out in the context of the ESF EuroBabel Programme. I would like to express my thanks to the ESF and AHRC for the funding received.
References


Bickel, Balthasar, Mao Raj, Netra P. Paudyal, Gona Banyade, Toya N. Bahatta, Marin Gaenzle, 
Elena Lieven, Idhtha Purma Raj, Novel Kishore Raj & Sabine Stoll. 2010. The syntax of 
three-argument verbs in Chintang and Belhare. In Andrej Malchukov, Martin Haspelmath & 
Bernard Comrie (eds.), *Studies in ditransitive constructions: A comparative handbook*, 


Bresnan Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative 
alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of 

British National Corpus (Oxford University 2000) http://www.bnc.ox.ac.uk.

Cardinaletti 2007. Subjects and wh-questions: Some new generalizations. In José Camacho, 
Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez & María José Cabrera (eds.), *Romance 
Linguistics 2006: Selected papers from the 36th Linguistic Symposium on Romance 
Languages (LSRL), New Brunswick, March–April 2006*, 57–78. Amsterdam & Philadelphia: 
John Benjamins.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency. In Henk van 
Gruyter.

Cattaneo, Abdrea 2009. *It is all about clitics: The case of the Northern Italian dialect 

Colleman, Timothy. 2009. Verb disposition in argument structure alternations: A corpus 


Gruyter.


Dziwierek, Katarzyna. 2002. A different kind of non-canonical case marking: The Slavic verb “to 

Linguistic Society* 16. 75–82.

glossos/issues/10/franks.pdf.

Gast, Volker. 2007. “I gave it him”: On the motivation for the alternative double object 


Appendix

Table 1: Number of DOC and Prep-C constructions identified for each verb and the occurrences of LP Rs in the BNC

<table>
<thead>
<tr>
<th>Verb</th>
<th>DOC</th>
<th>Prep-C</th>
<th>LP DOC</th>
<th>LP Prep-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>give</td>
<td>1,199</td>
<td>334</td>
<td>506</td>
<td>17</td>
</tr>
<tr>
<td>lend</td>
<td>283</td>
<td>368</td>
<td>147</td>
<td>21</td>
</tr>
<tr>
<td>hand</td>
<td>626</td>
<td>455</td>
<td>135</td>
<td>43</td>
</tr>
<tr>
<td>show</td>
<td>748</td>
<td>341</td>
<td>324</td>
<td>65</td>
</tr>
<tr>
<td>tell</td>
<td>1,837</td>
<td>28</td>
<td>1,183</td>
<td>12</td>
</tr>
<tr>
<td>ask</td>
<td>344</td>
<td>14</td>
<td>198</td>
<td>0</td>
</tr>
<tr>
<td>read</td>
<td>82</td>
<td>94</td>
<td>38</td>
<td>46</td>
</tr>
<tr>
<td>teach</td>
<td>282</td>
<td>94</td>
<td>134</td>
<td>5</td>
</tr>
<tr>
<td>send</td>
<td>139</td>
<td>201</td>
<td>82</td>
<td>13</td>
</tr>
<tr>
<td>forward</td>
<td>4</td>
<td>26</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>mail</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>bring</td>
<td>491</td>
<td>686</td>
<td>308</td>
<td>66</td>
</tr>
<tr>
<td>take</td>
<td>21</td>
<td>1,191</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>promise</td>
<td>80</td>
<td>4</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>offer</td>
<td>1,041</td>
<td>142</td>
<td>431</td>
<td>14</td>
</tr>
<tr>
<td>fling</td>
<td>10</td>
<td>17</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>throw</td>
<td>139</td>
<td>138</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>toss</td>
<td>25</td>
<td>25</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Total 7,354 4,160 3,550 313
Table 2: Number of DAC and Prep-C constructions identified for each verb and the occurrences of LP Rs in the CQO-web version of the PELCRA corpus

<table>
<thead>
<tr>
<th>Verb</th>
<th>DAC</th>
<th>Prep-C</th>
<th>LP DAC</th>
<th>LP Prep-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>dać</td>
<td>416</td>
<td>0</td>
<td>109</td>
<td>0</td>
</tr>
<tr>
<td>wręczyć</td>
<td>71</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>podać</td>
<td>118</td>
<td>23</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>pokazać</td>
<td>201</td>
<td>0</td>
<td>113</td>
<td>0</td>
</tr>
<tr>
<td>powiedzieć</td>
<td>941</td>
<td>41</td>
<td>662</td>
<td>8</td>
</tr>
<tr>
<td>przyznać</td>
<td>117</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>przekazać</td>
<td>373</td>
<td>60</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>uczyć</td>
<td>194</td>
<td>0</td>
<td>76</td>
<td>0</td>
</tr>
<tr>
<td>przysłać</td>
<td>65</td>
<td>56</td>
<td>37</td>
<td>7</td>
</tr>
<tr>
<td>przesłać</td>
<td>68</td>
<td>86</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>wysłać</td>
<td>28</td>
<td>303</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>przynieś</td>
<td>155</td>
<td>22</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>zanieś</td>
<td>8</td>
<td>30</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>zabrać</td>
<td>91</td>
<td>184</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>odebrać</td>
<td>173</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>obiecać</td>
<td>115</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>oferować</td>
<td>82</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>rzucić</td>
<td>22</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>3,238</td>
<td>810</td>
<td>1,208</td>
<td>24</td>
</tr>
</tbody>
</table>